Crystal Symmetry Algorithms in a High-Throughput Framework for Materials Research

by

Richard Taylor

Department of Mechanical Engineering and Materials Science
Duke University

Date: _____

Approved:

_____
Stefano Curtarolo, Supervisor

_____
Laurens Howle

_____
Nico Hotz

_____
Stefanos Venakides

Dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in the
Department of Mechanical Engineering and Materials Science
in the Graduate School of Duke University
2013

Abstract

Crystal Symmetry Algorithms in a High-Throughput Framework for Materials Research

by

Richard Taylor

Department of Mechanical Engineering and Materials Science
Duke University

Date: _____

Approved:

_____
Stefano Curtarolo, Supervisor

_____
Laurens Howle

_____
Nico Hotz

_____
Stefanos Venakides

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Mechanical Engineering and Materials Science in the Graduate School of Duke University
2013

Abstract

The high-throughput framework AFLOW that has been developed and used successfully over the last decade is improved to include fully-integrated software for crystallographic symmetry characterization. The standards used in the symmetry routines conform with the conventions and prescriptions given in the International Tables of Crystallography (ITC) (Hahn (2002)). A standard cell choice with standard origin is selected, and the space group, point group, Bravais lattice, crystal system, lattice system, and representative symmetry operations are determined. Following the conventions of the ITC, the Wyckoff sites are also determined and their labels and site symmetry are determined. The software makes no assumptions on the input cell orientation, origin, or lattice reduction and has been integrated in the AFLOW high-throughput framework for materials discovery by adding to the existing code base and making use of existing classes and functions. The software is written in object-oriented C++ for flexibility and reuse. A performance analysis and examination of the algorithms scaling with cell size and symmetry is also reported.

Contents

List of Tables

List of Figures

Acknowledgements

The author would like to thank Professor Stefano Curtarolo for his diligent mentoring and Shidong Wang, Kesong Yang, and Junkai Xue for fruitful discussions and support. The author would also like to extend special appreciation to Drs. Benjamin Burton, James Warren, and the entire Thermodynamics and Kinetics group at the National Institute of Standards and Technology (NIST) for their support and mentoring. This work was supported by NIST (AWARD No.70NANB12H163), ONR (N00014-11-1-0136,N00014-10-1-0436, N00014-09-1-0921), NSF (DMR-0639822,DMR-0908753), and the Department of Homeland Security Domestic Nuclear Detection Ofce.

1

Introduction

In recent years, computational materials science has greatly benefited from a high-throughput (HT) approach to materials search and discovery (Curtarolo et al. (2013)). System by system analyses lack the scope and speed needed to cover large portions of the materials landscape, whether they be experimental or computational by design. By performing broad searches over structural or chemical classes of materials and guided by fundamental physical principles, materials with specific desired properties can be systematically discovered and designed. The fundamental materials data that is the search space for HT materials design comprise quantum-mechanical and thermodynamic properties obtained using a range of reliable *ab initio* software tools (e.g., VASP (Kresse and Furthmuller (1996))). Ultimately, predicted materials should be validated (or invalidated) by experiment, and thereby, augment the knowledge base to construct more accurate databases and search models.

Critical to the HT materials design effort is the computational framework itself. The foundation that underpins this approach must be robust yet flexible software encompassing a wide range of tools for categorization and manipulation that integrate with a large materials properties database. The framework acts as "laboratory" for

computational-experimental design, execution and analysis. Accompanying all of this are the highly non-trivial tasks of input parameter control, job and memory management, database management, and the error handling that must accompany efficient HT computing.

In this dissertation I describe the addition of standardized crystallographic symmetry algorithms to AFLOW, a HT materials discovery framework that has been developed and successfully used over the last decade (S. Curtarolo, W. Setyawan, G. L. W. Hart, M. Jahnatek, R. V. Chepulskii, R. H. Taylor, S. Wang, J. Xue, K. Yang, O. Levy, M. Mehl, H. T. Stokes, D. O. Demchenko, and D. Morgan (2012); Curtarolo et al. (2012a)). The reason for a focus on crystallography is twofold, relating to current needs and future expectations: crystallography provides a theoretical basis by which the representation and characterization of physical materials is determined. The geometrical properties of materials lead to a rigorous scheme for materials description and bridge purely mathematical constructs to the physical characteristics of interest in applied materials science and engineering. Second, crystallography profoundly relates to real properties. It plays a fundamental role in determining electronic and thermal properties, magnetic properties, structural and mechanical properties, optical properties, etc. In short, crystal structure is intimately connected to physical properties because symmetry is intimately (and even existentially) connected to physical properties.

The crystallographic nature of all elemental materials and inorganic compounds in some region of phase space provides a starting point for materials science by group theoretical methods. Although the 230 crystallographic space groups (in three dimensions) have been nominally categorized for over one hundred years, the relationships between the full crystallographic characterization and the physical properties of materials provides impetus for ongoing research. The relatively recent discovery of quasi-crystals and incommensurately modulated structures and their connection

to higher-dimensional space groups is an important demonstration of a field that is still developing.

Many, if not all, of the computational tools in materials science rely on some level of symmetry characterization and reduction. Even imposing boundary conditions is an application of symmetry. As it relates to first-principles (ab initio) approaches, lattice periodicity more often becomes a fundamental assumption. Lattice periodicity allows the representation of an infinite crystal by a single unit cell. Without this assumption other constraints on the system must be made, and the problem may become intractable. One encounters the limits of current first-principles methods when the assumption of lattice periodicity is challenged (ie, by the introduction of crystal defects or boundaries). Additional symmetry information factors significantly into the implementation of computational tools for phonon spectra, elastic properties, energetics, and more.

Despite its fundamental importance as a building-block tool in computational materials science, a few algorithms for the full *standard* characterization of crystal symmetry have been developed and implemented (Stokes et al. (2013)). Furthermore, little attention has been payed to the numerical cost of the developed algorithms — a question of increasing significance as high-throughput approaches are transforming the approach toward materials discovery. As high-throughput materials discovery is pushed to its limits, faster, more efficient discovery algorithms will be needed. Finally, and most importantly, to fully realize the potential of HT materials design, functional components of the computational framework must be fully integrated for extensibility, reuse, and efficiency (ie, it is not sufficient to simply use the output of an isolated code).

In this dissertation, I describe the methods and implementation of software for full space group determination and characterization that conforms with the standards of the International Tables of Crystallography (Hahn (2002)). A standard cell

choice with standard origin is selected, and the space group, point group, Bravais lattice, crystal system, lattice system, and representative symmetry operation set are determined. The algorithm makes no assumptions on the input cell orientation, origin, or reduction and has been implemented using the main classes and libraries of the AFLOW high-throughput framework for materials discovery. It is available in AFLOW and AFLOW-online as a fully integrated tool.

Beyond the determination of the crystal space group, a routine for finding the standard crystal Wyckoff sites is also implemented. Knowledge of the Wyckoff positions allows for an additional level in the hierarchy of crystal characterization, by standardizing labels for the symmetrically unique groups of sites within the crystal. This allows for more refined crystallographic analysis and provides the information for accelerating related routines (crystal distortions, phase transitions, etc). A basic numerical analysis of the algorithm's performance as a function of symmetry and cell size is given with a discussion of the various instances of tolerance and tolerance control included in the algorithms.

Because a clear understanding of fundamental crystallographic concepts and conventions is crucial to the understanding of the symmetry algorithm, I devote a portion of this dissertation to their explanation (Chapter 2). The algorithm itself is described in detail in Chapters 3 and 4—with Chapter 4 being devoted to a detailed functional description of the routines. A conclusion is given in Chapter 5.

2

Symmetry in Crystallography: Crystal System, Point Group, and Space Group

The symmetry properties of crystals provide the foundation for the categorization and study of materials. These properties are divided by several classification schemes which include the *crystal system* (or crystal family), *lattice system*, *point group*, and *space group*. The physical properties of crystals are intimately connected with these symmetry descriptors. For example, crystals belonging to a centrosymmetric space group cannot exhibit the piezoelectric effect due to the existence of an inversion center. Everything from electronic structure and optical properties to lattice vibrations are profoundly related to the crystal symmetry. In an effort to study the properties of materials it is, therefore, necessary to have tools for the accurate classification and description of the symmetry relations.

2.1 Proper rotations, inversions, and reflections

The assumption that crystals exhibit translational symmetry limits possible symmetries to orders of two, three, four, and six. Rotation is defined in terms of its order ($2^{\pm}$,$3^{\pm}$,$4^{\pm}$,or $6^{\pm}$) and an axis, about which the rotation occurs (e.g. 3^{+} x, x, x

indicates a threefold *counter-clockwise* rotation about the line x, x, x). Inversion symmetry is reflection about a point (e.g., about the origin, $(x, y, z) \rightarrow (-x, -y, -z)$) and is uniquely defined by the inversion center. Finally, a reflection is the operation of mirroring through a plane and is uniquely defined by the plane itself (in two dimensions, reflection is equivalent to two-fold rotation; in general, they differ in $n > 2$ dimensions).

2.2 Compound symmetry

The compound symmetry operations include combinations of rotations, reflections, inversions, and internal translations (Table 2.1). Combined with the point operations and the translational symmetry of the lattice, these define the space group.

Table 2.1: Compound symmetry operations.

	Rotation	**Reflection**	**Inversion**	**Translation**
Rotation	Rotation	Rotoreflection	Rotoinversion	Screw rotation
Reflection	—	Reflection	2-fold rotation	Glide reflection
Inversion	—	—	Inversion	Inversion
Translation	—	—	—	Translation

2.3 Crystal class or point group and Hermann-Mauguin notation

There are thirty-two crystal classes or crystallographic point groups. These comprise the possible groups of vector operations that leave the external morphology of crystals invariant. For general objects not constrained by translational symmetry, there are infinitely many such groups. However, in crystals these are reduced by the requirement of translational symmetry which does not permit five-fold and $(n > 6)$-fold rotation symmetry. Each of the thirty-two crystal classes can be uniquely represented by a geometric, space-filling object called a *crystal form*.

It is important to note that, within the context of space group symmetry elements, it is only the *character* of the operation that defines the point group, and so operations within a crystal point group may not, in general, be isometries of the

crystal. As stated above, the crystal point group is the group of isometries acting on the *vector set of face normals* of the macroscopic crystal. It is a property of the external morphology of the crystal and is uniquely determined by the crystal space group.

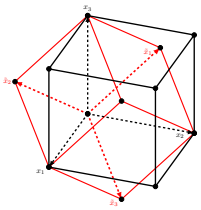
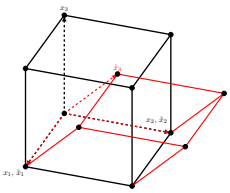
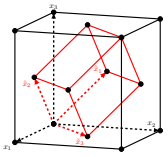
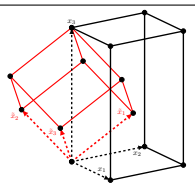
The notation used throughout this dissertation and in the International Tables is the so-called Hermann-Mauguin notation, perhaps the most commonly used in crystallography. Rotation axes are denoted by their rotation order (i.e., 1, 2, 3, 4, 6 corresponding to rotations of 2π, π, $\pi/2$, and so forth). A '$-$' is used to indicate an improper rotation (e.g., $\bar{2}$ is a two-fold rotation-inversion). Mirror planes—which are equivalently described as a two-fold rotation-inversion—are denoted by m. When a rotation axis is parallel to the normal of a mirror plane, both are indicated by a fraction (e.g., $2/m$). The Hermann-Mauguin notation for point groups does not explicitly list the complete set of point operations. Instead, only those symmetries along special directions are given. These are the *primary*, *secondary*, and *tertiary* directions and are defined within a crystal system.

2.4 Lattice system and crystal system

2.4.1 Lattice system and Bravais lattices

The crystal lattice is the set of points defined by the span of three linearly independent vectors. The representation of the lattice by a vector basis is not unique, and the lattice vectors may vary in both orientation, length, and integer-multiple changes of volume (with appropriate changes to the atomic basis). This ambiguity, however, is only superficial as lattice reduction and examination of the symmetry of the lattice can be used to specify a unique description. In general, one considers lattice representation of two types: conventional and primitive. The primitive representation is specified by metric conditions on the lattice vectors and differs from the conventional cell only when a centering operation exists (see Table 2.2.

Table 2.2: Standard conventional to standard primitive transform for the centered Bravais lattices.

Orthorhombic/ Tetragonal/ Cubic I	$\begin{pmatrix}\tilde{x}_1\\ \tilde{x}_2\\ \tilde{x}_3\end{pmatrix} = \begin{pmatrix}-\frac{1}{2} & \frac{1}{2} & \frac{1}{2}\\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2}\\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2}\end{pmatrix}\begin{pmatrix}x_1\\ x_2\\ x_3\end{pmatrix}$	
Monoclinic/ Orthorhombic C	$\begin{pmatrix}\tilde{x}_1\\ \tilde{x}_2\\ \tilde{x}_3\end{pmatrix} = \begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & \frac{1}{2} & \frac{1}{2}\end{pmatrix}\begin{pmatrix}x_1\\ x_2\\ x_3\end{pmatrix}$	
Orthorhombic/ Cubic F	$\begin{pmatrix}\tilde{x}_1\\ \tilde{x}_2\\ \tilde{x}_3\end{pmatrix} = \begin{pmatrix}0 & \frac{1}{2} & \frac{1}{2}\\ \frac{1}{2} & 0 & \frac{1}{2}\\ \frac{1}{2} & \frac{1}{2} & 0\end{pmatrix}\begin{pmatrix}x_1\\ x_2\\ x_3\end{pmatrix}$	
Hexagonal R	$\begin{pmatrix}\tilde{x}_1\\ \tilde{x}_2\\ \tilde{x}_3\end{pmatrix} = \begin{pmatrix}\frac{2}{3} & \frac{1}{3} & \frac{1}{3}\\ -\frac{1}{3} & \frac{1}{3} & \frac{1}{3}\\ -\frac{1}{3} & -\frac{2}{3} & \frac{1}{3}\end{pmatrix}\begin{pmatrix}x_1\\ x_2\\ x_3\end{pmatrix}$	

The *lattice system* of crystals forms a classification scheme of space groups based on groupings of the possible point groups of crystallographic lattices. In three dimensions there are seven lattice systems: cubic, tetragonal, hexagonal, rhombohedral, orthorhombic, monoclinic, triclinic (Fig. 2.2). The addition of *centering translations* which preserve the lattice point symmetry but alter the translational symmetry, results in the fourteen so-called Bravias lattices.

2.4.2 Crystal system

The *crystal systems* are nominally similar to the lattice systems: There are seven crystal systems, five of which are named identically to their lattice-system counterparts. The discrepancy arises in the trigonal and hexagonal crystal systems (the trigonal crystal system includes space groups belonging to both the rhombohedral and hexagonal lattice systems), and is due to underlying differences inherent in these

modes of classification: The crystal system is defined by the point group of the crystal but not by its lattice. For example, crystals containing a 3-fold symmetry axis belong to the trigonal crystal system. However, such crystals may have either a rhombohedral or hexagonal lattice.

Some of the confusion this causes may be averted by considering instead six *crystal families* and the equivalent six *lattice families*. In these schemes, the trigonal and hexagonal crystal systems are grouped into one hexagonal crystal family, and the rhombohedral and hexagonal lattice systems into one hexagonal lattice family (Fig. 2.1).

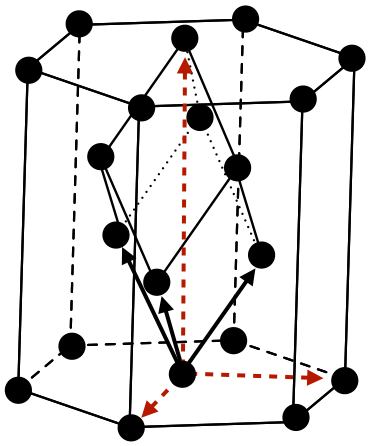


FIGURE 2.1: Ambiguity in the choice of unit cell is shown. A rhombohedral unit cell (solid lattice vectors) or the larger hexagonal cell (dashed lattice vectors) generate the same lattice.

The crystal system of a space group is derived from its crystal class by identifying key symmetry operations described in Table 3.4. The overlap between the hexagonal and rhombohedral lattice systems occurs for those space groups having 3, 3_1, 3_2 or $\bar{3}$ as principal axis. These comprise the *trigonal* crystal system but may be in either the *hexagonal* or *rhombohedral* lattice system.

2.5 Space group

There are 230 unique groupings of the crystallographic point operations, lattice translation symmetries, and compound symmetry operations (glide, screw, etc). Combining the 14 Bravais lattices with the point operations gives the 73 *simple space groups*. The simple space groups are also called the *arithmetic crystal classes*. The addition of glide planes and screw axes brings the total number of *affine space groups* to 219. The 219 space groups are classified by their underlying abstract groups: two space groups are equivalent if an affine transformation exists relating one to the other. In practice, we want to distinguish between right and left handed screw operations as well as preserving the handedness of the coordinate system. These conditions expand the space group types to the 230 *crystallographic space groups* (Table 2.3).

As a final note on the space groups, the *holosymmetric space groups* are those constructed by only applying the possible lattice translations to a point. In these cases, the point group of the space group is equal to the point group of the lattice. The seven holohedral point groups are $\bar{1}$, $2/m$, mmm, $4/mmm$, $\bar{3}m$, $6/mmm$, and $m\bar{3}m$. The relation between these symmetry categories is given in Figure 2.3.

Table 2.3: The 230 space groups with point groups and crystal systems. The space groups and point groups are given using Hermann-Mauguin notation. The Laue groups indicating the centrosymmetric point groups are boxed.

Crystal System	Point Group	Space Groups
Triclinic (1-2)	1	$P1$
	$\boxed{\bar{1}}$	$P\bar{1}$
Monoclinic (3-15)	2	$P2$, $P2_1$, $C2$
	m	Pm, Pc, Cm, Cc
	$\boxed{2/m}$	$P2/m$, $P2_1/m$, $C2/m$, $P2/c$, $P2_1/c$, $C2/c$
Orthorhombic (16-74)	222	$P222$, $P222_1$, $P2_12_12$, $P2_12_12_1$, $C222_1$, $C222$, $F222$, $I222$, $I2_12_12_1$
	$mm2$	$Pmm2$, $Pmc2_1$, $Pcc2$, $Pma2$, $Pca2_1$, $Pnc2$, $Pmn2_1$, $Pba2$, $Pna2_1$, $Pnn2$, $Cmm2$, $Cmc2_1$, $Ccc2$, $Amm2$, $Aem2$, $Ama2$,

	$\boxed{mmm}$	$Aea2$, $Fmm2$, $Fdd2$, $Imm2$, $Iba2$, $Ima2$ $Pmmm$, $Pnnn$, $Pccm$, $Pban$, $Pmma$, $Pnna$, $Pmna$, $Pcca$, $Pbam$, $Pccn$, $Pbcm$, $Pnnm$, $Pmmn$, $Pbcn$, $Pbca$, $Pnma$, $Cmcm$, $Cmce$, $Cmmm$, $Cccm$, $Cmme$, $Ccce$, $Fmmm$, $Fddd$, $Immm$, $Ibam$, $Ibca$, $Imma$
Tetragonal (75-142)	4	$P4$, $P4_1$, $P4_2$, $P4_3$, $I4$, $I4_1$
	$\bar{4}$	$P\bar{4}$,$I\bar{4}$
	$\boxed{4/m}$	$P4/m$, $P4_2/m$, $P4/n$, $P4_2/n$, $I4/m$, $I4_1/a$
	422	$P422$, $P42_12$, $P4_122$, $P4_12_12$, $P4_222$, $P4_22_12$, $P4_322$, $P4_32_12$, $I422$, $I4_122$
	$4mm$	$P4mm$, $P4bm$, $P4_2cm$, $P4_2nm$, $P4cc$, $P4nc$, $P4_2mc$, $P4_2bc$, $I4mm$, $I4cm$, $I4_1md$, $I4_1cd$
	$\bar{4}2m$	$P\bar{4}2m$, $P\bar{4}2c$, $P\bar{4}2_1m$, $P\bar{4}2_1c$, $P\bar{4}m2$, $P\bar{4}c2$, $P\bar{4}b2$, $P\bar{4}n2$, $I\bar{4}m2$, $I\bar{4}c2$, $I\bar{4}2m$, $I\bar{4}2d$
	$\boxed{4/mmm}$	$P4/mmm$, $P4/mcc$, $P4/nbm$, $P4/nnc$, $P4/mbm$, $P4/mnc$, $P4/nmm$, $P4/ncc$, $P42/mmc$,$P42/mcm$, $P4_2/nbc$, $P4_2/nnm$, $P4_2/mbc$, $P4_2/mnm$, $P4_2/nmc$, $P4_2/ncm$, $I4/mmm$, $I4/mcm$, $I4_1/amd$, $I4_1/acd$
Trigonal (143-167)	3	$P3$, $P3_1$, $P3_2$, $R3$
	$\boxed{\bar{3}}$	$P\bar{3}$, $R\bar{3}$
	32	$P312$, $P321$, $P3_112$, $P3_121$, $P3_212$, $P3_221$, $R32$
	$3m$	$P3m1$, $P31m$, $P3c1$, $P31c$, $R3m$, $R3c$
	$\boxed{\bar{3}m}$	$P\bar{3}1m$, $P\bar{3}1c$, $P\bar{3}m1$, $P\bar{3}c1$, $R\bar{3}m$, $R\bar{3}c$
Hexagonal (168-194)	6	$P6$, $P6_1$, $P6_5$, $P6_2$, $P6_4$, $P6_3$
	$\bar{6}$	$P\bar{6}$
	$\boxed{6/m}$	$P6/m$, $P6_3/m$
	622	$P622$, $P6_122$, $P6_522$, $P6_222$, $P6_422$, $P6_322$
	$6mm$	$P6mm$, $P6cc$, $P6_3cm$, $P6_3mc$
	$\bar{6}m2$	$P\bar{6}m2$, $P\bar{6}c2$, $P\bar{6}2m$, $P\bar{6}2c$
	$\boxed{6/mmm}$	$P6/mmm$, $P6/mcc$, $P6_3/mcm$, $P6_3/mmc$
Cubic (195-230	23	$P23$, $F23$, $I23$, $P2_13$, $I2_13$
	$\boxed{m\bar{3}}$	$Pm\bar{3}$, $Pn\bar{3}$, $Fm\bar{3}$, $Fd\bar{3}$, $Im\bar{3}$, $Pa\bar{3}$, $Ia\bar{3}$
	432	$P432$, $P4_232$, $F432$, $F4_132$, $I432$, $P4_332$,

		$P4_132$, $I4_132$
	$\bar{4}3m$	$P\bar{4}3m$, $F\bar{4}3m$, $I\bar{4}3m$, $P\bar{4}3n$, $F\bar{4}3c$, $I\bar{4}3d$
	$\boxed{m\bar{3}m}$	$Pm\bar{3}m$, $Pn\bar{3}n$, $Pm\bar{3}n$, $Pn\bar{3}m$, $Fm\bar{3}m$, $Fm\bar{3}c$, $Fd\bar{3}m$, $Fd\bar{3}c$, $Im\bar{3}m$, $Ia\bar{3}d$

2.6 Wyckoff sites and their evaluation

The positions of atoms and their site symmetries play a central role in the description of a crystal structure. The space group alone does not provide information regarding the number of atoms within a crystallographic cell (the stoichiometry of the structure) or explicitly distinguish between atoms of equivalent types and the symmetry relations between them. In general, when referring to atomic positions in a crystal one speaks of *general positions* and *special positions*. The general position is the point set containing (x, y, z) such that the no operation in the set of group operations $\mathcal{G}$ leaves the point fixed. A special position is one in which at least one operation in $\mathcal{G}$ leaves the point fixed.

A method by which the Wyckoff positions can be derived is shown with the following simple example.

Wyckoff Position Derivation

Consider the orthorhombic space group #34-$Pnn2$ with symmetry operations comprising the identity, two glide planes and a 2-fold rotation about the z axis: (1) 1 (2) 2 0 0 z (3) n(1/2,0,1/2) x 1/4 z and (4) n(0,1/2,1/2) 1/4 y z. An arbitrary position (x, y, z) has the following equivalent positions (applying the operations (1)-(4) in the order listed): (1) (x,y,z) (2) $(-x,-y,z)$ (3) $(x + 1/2,-y + 1/2,z + 1/2)$ and (4) $(-x + 1/2,y + 1/2,z + 1/2)$. This is the general position, as there are no operations which leave the point fixed, when no assumptions are made on the values of x, y, and z . To arrive at the special positions, let us systematically impose equivalencies:

(1) = (2):

$$x = -x \Rightarrow x = 0 \tag{2.1}$$

$$y = -y \Rightarrow y = 0 \tag{2.2}$$

$$z = z \tag{2.3}$$

Substituting $x = y = 0$ into (1)-(4) gives the special position: (1′) $(0, 0, z)$ (2′) $(1/2, 1/2, z + 1/2)$. In the International Tables, this is designated as position 2 *a* ..2 (multiplicity, Wyckoff letter, and site symmetry). Because of translational symmetry an additional equivalence can be found by solving (1) = (2) modulo 1.

(1) = (4):

$$x = -x \Rightarrow x = 1/2 \tag{2.4}$$

$$y = -y \Rightarrow y = 1/2 \tag{2.5}$$

$$z = z \tag{2.6}$$

Substituting $x = y = 1/2$ into (1)-(4) gives the special position: (1′) $(1/2, 1/2, z)$ (2′) $(0, 0, z + 1/2)$. If we let $z \rightarrow z + 1/2$ and reduce modulo one, the position is represented as position 2 *b* ..2 in the International Tables.

2.7 C++ symmetry classes in the code

The search for point symmetries and comparison with the International space group tables is greatly facilitated by the construction of general symmetry classes described below. The flexibility of the class construction within an object oriented environment will also allow for future extensions and adaptations of the code. The space group comparison routine itself makes use of a library of symmetry operations containing the information from the 230 space groups and the possible point operations divided by crystal class.

2.7.1 Improper rotations and the `RotoInversion` *class*

Rotation-inversion (rotoinversion) symmetries (rotation about an axis followed by inversion about a point on the axes is equivalent to rotation about the axis followed by reflection in a plane perpendicular to the axis.) (Fig. 2.4) are compound symmetry operations that combine proper rotations with inversion about a point on the rotation axis. The possible rotoinversions in crystallography are one-fold, two-fold, three-fold, four-fold, and six-fold. One-fold rotoinversion is equivalent to a simple inversion. Two-fold rotoinversion is equivalent to a reflection symmetry with mirror plane orthogonal to the rotation axis. Three-fold rotoinversion occurs in crystals with three-fold rotation and inversion (the inversion may coincide with the standard origin). Six-fold rotoinversion can be identified by a three-fold rotation and a mirror plane orthogonal to the rotation axis. Four-fold rotoinversion cannot be equivalently described by simpler operations.

Rotoinversion is the most general of the point operations (Fig. 3.5): Pure rotations, reflections, and inversions can be described as types of rotoinversion. Thus, the `RotoInversion` class can be applied to all the point group operations. Rotation angle, rotation axis, and inversion point uniquely define the rotoinversion. In the `RotoInversion` class, these are implemented as private members. The class variables are filled via the `void get_roto_inversion(string s)` function. The string `s` takes a standard format for rotoinversions (similar to that used in the ITC). For example, a three-fold counterclockwise rotation about an axes defined parametrically, $(x, x, 0)$, and an inversion point triple, $(0, 0, 0)$, is described by 3^+ x x 0; 0 0 0.

The rotoinversion operation is applied using the matrix representations of the rotation and inversion operations. The rotation matrix is uniquely defined by the specification of the angle α and the rotation axis. The axis is defined by a direction vector (u, v, w) and a point (a, b, c) on the axis. Using this notation, the matrix is,

$$\begin{bmatrix} u^2 + x\cos(\alpha) & uv(1-\cos(\alpha)) - w\sin(\alpha) \\ uv(1-\cos(\alpha)) + w\sin(\alpha) & v^2 + y\cos(\alpha) \\ uw(1-\cos(\alpha)) - v\sin(\alpha) & vw(1-\cos(\alpha)) + u\sin(\alpha) \\ 0 & 0 \end{bmatrix} \tag{2.7}$$

$$\begin{bmatrix} uw(1-\cos(\alpha)) + v\sin(\alpha) & (ax - u(bv+cw))(1-\cos(\alpha)) + (bw-cv)\sin(\alpha) \\ vw(1-\cos(\alpha)) - u\sin(\alpha) & (by - v(au+cw))(1-\cos(\alpha)) + (cu-aw)\sin(\alpha) \\ w^2 + z\cos(\alpha) & (cz - w(au+bv))(1-\cos(\alpha)) + (av-bu)\sin(\alpha) \\ 0 & 1 \end{bmatrix}$$

To slightly simplify the unwieldy form, a few substitutions have been made:

$$v^2 + w^2 \equiv x \tag{2.8}$$

$$u^2 + w^2 \equiv y \tag{2.9}$$

$$u^2 + v^2 \equiv z \tag{2.10}$$

Following application of the 4x4 rotation matrix—a general rotation about a line is an affine transformation and so requires an additional row and column to account for the "translation part"—the inversion about a point (x_0, y_0, z_0) is given by,

$$\begin{bmatrix} -1 & 0 & 0 & 2x_0 \\ 0 & -1 & 0 & 2y_0 \\ 0 & 0 & -1 & 2z_0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2.11}$$

In each of these cases, the operand is a vector with four components. The first three are the standard coordinates of a point and the fourth component is always unity.

2.7.2 Glide class

The glide operation is defined as a reflection followed by a translation (or equivalently, the same in reverse order). The `Glide` class in filled using the `void get_glide(string str)` function. The input string is in a form similar to

that used in the ITC, with abbreviated notation for *axial, diagonal, and diamond* glide operations given by $\{a, b, c\}, n, d$.

Because the reflection plane may not pass through the origin, the operation includes reflection and translation parts. The reflection component uses the plane normal (n_1, n_2, n_3) and is represented by the 3x3 matrix,

$$\begin{bmatrix} n_2^2 + n_3^2 & -n_1 n_2 & -n_1 n_3 \\ -n_1 n_2 & n_1^2 + n_3^2 & -n3n_2 \\ -n_3 n_1 & -n_3 n_2 & n_1^2 + n_2^2 \end{bmatrix}. \tag{2.12}$$

The full operation on P is

$$\mathbf{A}\vec{P} + d\vec{n} + |\vec{n} \cdot \vec{P} - d|\vec{n} + \vec{T}, \tag{2.13}$$

where d is defined by the standard form of the plane, $ax + by + cz + d = 0$ (the triple (a, b, c) is parallel to the plane normal). Care must be given to the orientation of the selected normal. The normal should direct away from the operand, P. The `Glide` class corrects an incorrectly selected normal. T is the glide translation.

2.7.3 Screw class

A screw axis (Fig. 2.6) is a compound symmetry operations consisting of a rotation followed by a translation parallel to the rotation axis. Screw axes are implemented in the symmetry algorithm code with the `Screw` class and the member function `void get_screw(string str)`. The input string follows the standard in the ITC. For example, a twofold screw axis about $x, x, 1/4$ with translation $(1/2, 1/2, 0)$ is described by the string 2 (1/2 1/2 0) x x 1/4. The rotation matrix is exactly that described for the rotation-inversion class (2.7).

2.7.4 Operator overloading in the symmetry classes

The ease with which the various symmetry classes can be used in the developed routines and algorithms greatly depends on the simplicity of their application to

point triplets. Although the use of the classes here is solely for the determination the crystallographic space group (via a particular algorithmic approach), their usefulness extends beyond this. They can, for example, be used for more general symmetry applications or for alternative space group determination algorithms (e.g., by complete determination of the space group symmetries). To facilitate their use and portability the symmetry types exist as distinct classes and are applied to a point by overloading the $*$ operator. This allows the user to work with the (often complex) symmetry classes in a truly object-oriented fashion with limited concern for the nature of internal data types and functions.

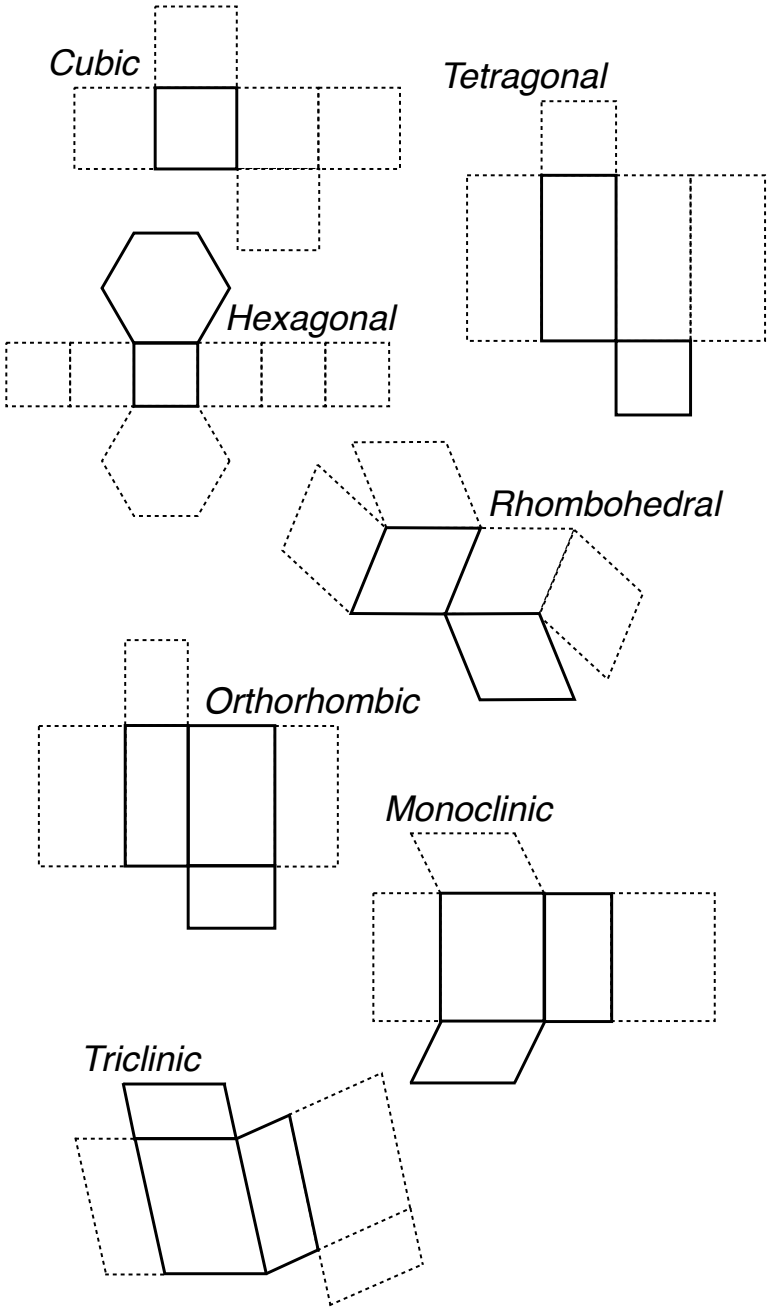


FIGURE 2.2: The seven lattice systems in 3 dimensions with unfolded faces.

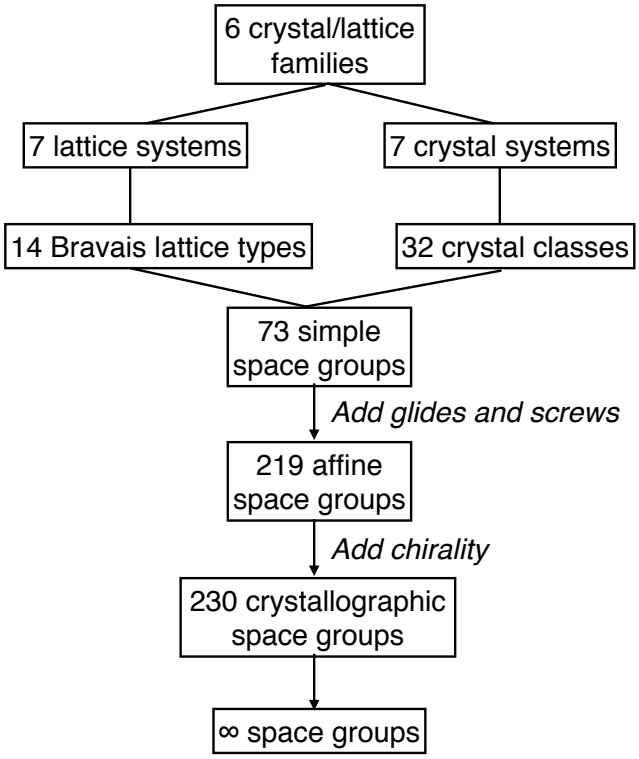


FIGURE 2.3: Classification of crystal structures. Based on Fig. 8.2.1.1 in ITC (Hahn (2002))

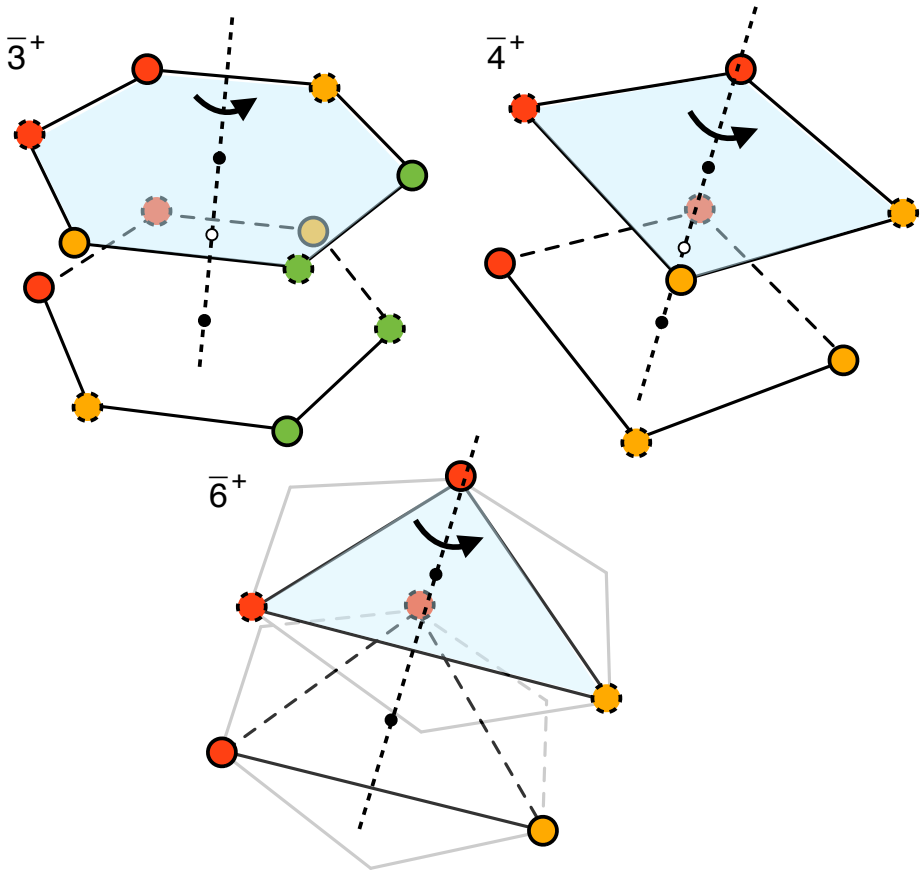


FIGURE 2.4: Three-fold, four-fold, and six-fold rotoinversions are illustrated by successive application to a atomic position. The sites shown with a dashed circle indicate a projection of an atomic position into a plane of equivalent atoms under a pure rotation. In all but the six-fold case, the dashed sites do not represent generated positions, but rather show the relation to the simple rotational case.

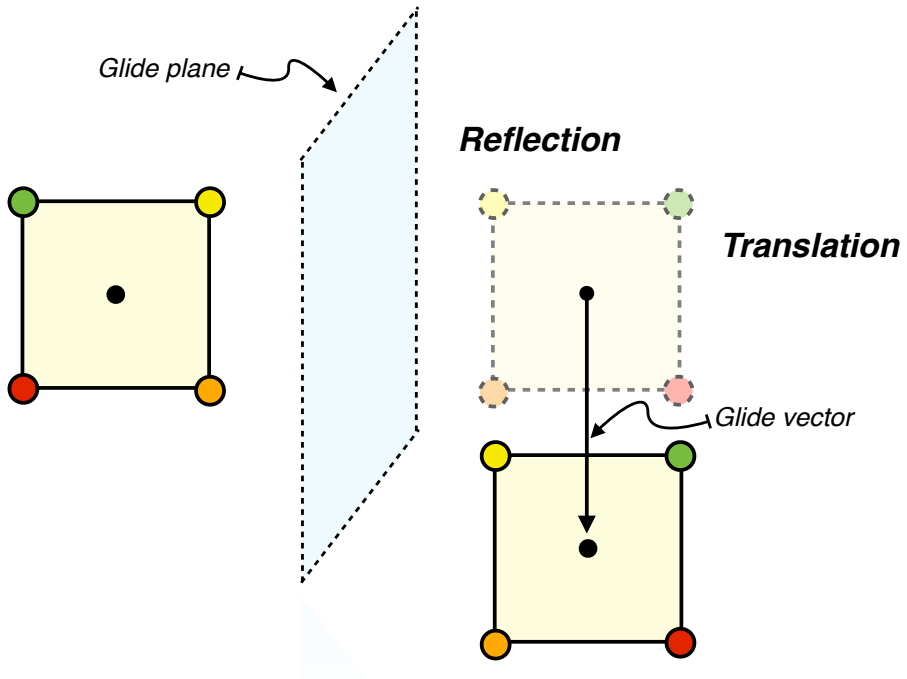


FIGURE 2.5: Cartoon illustrating a glide operation.

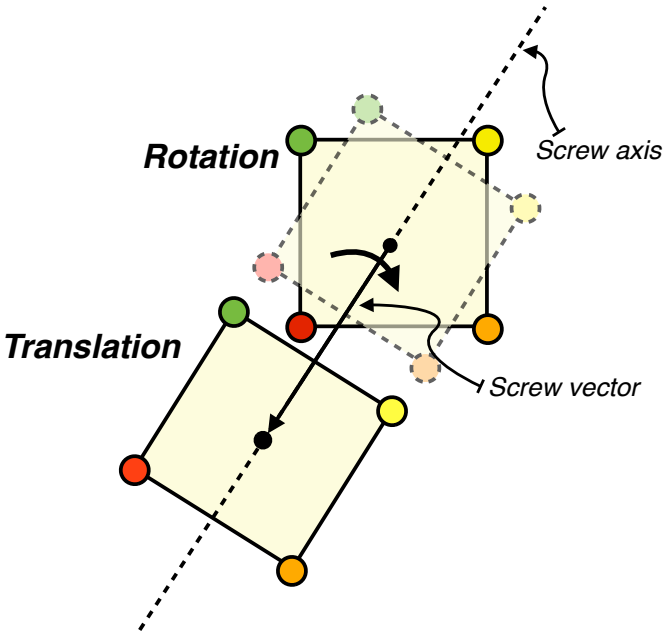


FIGURE 2.6: Cartoon illustrating a screw operation.

3

Overview of the symmetry routines and related functions

3.1 Overview of the algorithm

Ab initio research of material properties relies first on an accurate structural description of the material(s) of interest. In general, only the most basic structural information (lattice and atomic basis) is known. Although symmetry properties such as the space group and the Wyckoff positions are implied by the structure file (e.g., POSCAR (Kresse and Hafner (1993))), to make use of these, one requires an explicit representation. As a structure representation explicitly containing the Wyckoff positions, the analog to the POSCAR is the so-called WYCCAR (S. Curtarolo, W. Setyawan, G. L. W. Hart, M. Jahnatek, R. V. Chepulskii, R. H. Taylor, S. Wang, J. Xue, K. Yang, O. Levy, M. Mehl, H. T. Stokes, D. O. Demchenko, and D. Morgan (2012)). In the context of high-throughput (HT) computations and analyses, there is a need for extraction of the implicit symmetry information via some algorithmic process. It is possible, however tedious, to determine the symmetry relations by inspection or through the use of an externally called or internally wrapped program. This is an unpractical and unsatisfactory method when hundreds

or thousands of structures must be analyzed and when the process is to be part of a tailored, automated optimization, search, or comparison scheme. Furthermore, it does not allow for future modification or augmentation of the routines.

The algorithm presented here relies principally on the data provided in the fifth edition of the International Tables of Crystallography (ITC) (Hahn (2002)). The overarching methodology is simple in concept, although due to the intricacies of crystallographic representation and symmetry, it is non-trivial to implement in practice. The routine requires now that a standard POSCAR format is provided as input (S. Curtarolo, W. Setyawan, G. L. W. Hart, M. Jahnatek, R. V. Chepulskii, R. H. Taylor, S. Wang, J. Xue, K. Yang, O. Levy, M. Mehl, H. T. Stokes, D. O. Demchenko, and D. Morgan (2012)), although the code could easily be adapted to accept alternate forms of crystal structure input. In fact, such flexibility has already been added to the underlying crystal structure class within AFLOW. The basic algorithm is essentially the sequential implementation of four conceptually distinct parts (Fig. 3.1): a) Find the conventional-cell representation of the crystal by a partial symmetry search; b) Use reduced library of symmetry operations to determine point group and space group operations; c) Compare combinations of symmetry set to library of generators to find standard origin shift and space group number; d) Solve for Wyckoff site variables using library and set of equivalent atoms.

Not surprisingly, this overview of the algorithm overlooks details critical to the actual implementation of the algorithm, however, it should suffice as a 'big picture' view of the process. As one level of additional detail, the process of lattice reduction, evaluating the lattice symmetry, a loop for consistency of the conventional cell representation, and reduction of the space group candidate set by prior information are also included (Fig. 3.1). A detailed function by function overview of the algorithm is given in Chapter 4.

3.2 Lattice symmetry and the conventional cell

A general strategy employed in the design of the space group determination algorithm was the use of standard reference tables and libraries whenever possible. This is desirable from a computational standpoint as it significantly reduces needless combinatorial searches and integrates the end goal of producing standardized output into the algorithm itself. However, without any prior knowledge of the lattice basis and cell orientation, reference to standard tables is of no use. The representation of the symmetry operations and Wyckoff sites are both cell-choice and origin dependent.

To make use of the symmetry libraries for point operations and space groups in the ITC, the standard conventional lattice vectors must first be found, and the crystal must be transformed accordingly. The approach is to use a limited set of the crystal symmetry, namely the crystal class, and then, iterate a process which reveals the symmetry operations with reference to the origin. The critical point is that the lattice symmetry implies either the correct crystal class or one which can be iterated to the correct crystal class by comparison of the crystal class implied by the full symmetry set and that suggested by the lattice. Once the crystal class and origin-specific symmetries are known, the standard conventional cell can be derived. The use of specific symmetry directions of the crystal to define a conventional cell is described in Section 3.4.

The conventional cell routine relies on the fact that the class of the crystal will be equal to or a subgroup of the class implied by the lattice symmetry (Fig. 3.13). This works to give a correct conventional cell in at most two iterations, excluding the iterations necessary to account for the ambiguity that arises when equivalent symmetry directions are used to specify unit cell vectors. At most two iterations are required because the full symmetry set can be computed after the initial conventional cell is constructed, using the library of point operations. When a cell ambiguity exists

(such as in the orthorhombic space groups), unique permutations of the ambiguous directions are attempted until a match is found. Similar to the point operations listed in the ITC, the library of point operations in the space group code is divided according to cubic, tetragonal, orthorhombic, monoclinic, triclinic, or rhombohedral coordinate systems, and into the hexagonal coordinate system. Referring to the group-subgroup relations, the trigonal point groups are subgroups of the cubic group, and thus, for example, the complication that may exists for the trigonal space groups represented with a cubic lattice is easily handled.

3.3 Method for finding lattice symmetry of arbitrary crystal structure

The algorithm for computing the lattice symmetry directions (for the zeroth iteration) makes use of the geometric properties underlying the crystallographic symmetries. Due to the requirement of translation symmetry in traditional crystals, the order of rotations considered are limited to $n = 1, 2, 3, 4, 6$. Rotation symmetries of other orders exist only in quasi-crystalline structures. In addition to the rotation isometries, there exist inversions, plane reflections and improper rotations (an n-fold rotation followed by inversion about a point). Improper rotations are termed "rotoinversions" or "rotoreflections" and are indicated by their order with a macron (e.g., a threefold rotoinversion would be $\bar{3}$). The geometric structures related to the possible point symmetries of crystals are *line, plane, equilateral triangle, isosceles right triangle* and *120-30-30 triangle* in connection with $n = 1, 2, 3, 4, 6$ respectively.

A search over all sets of co-planar atoms (sets of three) for specific geometric relations (Fig. 3.4 and 3.5) reveals point operations of the lattice. As an example, consider a search for mirror planes and threefold rotation axes (Fig. 3.3). Mirror planes must exist at the midpoint of two atoms with normal defined by the connecting line. Therefore, a search over all unique pairs of atoms with the appropriately defined candidate mirror planes will yield the plane symmetries. Similarly, a three-

fold rotation axes must exist at the center point of three atoms forming an equilateral triangle. The rotation axes must be normal to the plane defined by the three atoms. Searching all unique sets of three atoms, checking for an equidistant center of mass and then applying the candidate rotations to the unit cell will yield the threefold symmetries.

3.4 Standard Conventional Unit Cell

When confronted with the problem of selecting a lattice unit cell, lattice primitive reduction is often the chosen method, as it is generally the most straightforward to implement. A reduced lattice basis refers to one with the smallest possible volume. The minimal volume condition does not imply a unique cell, however (see Fig. 3.6). While there are a number of difficulties to the reduced-lattice approach (including implementing a stable algorithm) (Wübben et al. (2011); Grosse-Kunstleve et al. (2004)), the reduced cell is important in a number of crystallographic applications (e.g, crystallographic classification and categorization). In the 1920s, Niggli, drawing on the work of Eisenstein, defined algebraic conditions leading to a unique, reduced lattice basis (Niggli (1928)). The so-called Niggli reduced cell is frequently used; a Niggli reduction algorithm is implemented in AFLOW (S. Curtarolo, W. Setyawan, G. L. W. Hart, M. Jahnatek, R. V. Chepulskii, R. H. Taylor, S. Wang, J. Xue, K. Yang, O. Levy, M. Mehl, H. T. Stokes, D. O. Demchenko, and D. Morgan (2012)).

Reduction procedures such as the Niggli approach cited above rely on metric relations (i.e., relations restricting angles between lattice vectors and their lengths) to determine the unit cell. In contrast to a metric approach, it is also possible to use the symmetry properties of the lattice to determine the unit cell. Use of symmetry in this way leads to the conventional unit cell. This has the advantage of being consistent with the overall goal of the symmetry algorithm—determination of point

group and space group. The conventional cell defined in this manner also serves as a more insightful visual representation of the structure, due to its relation to the structure's symmetry.

The conventional cell constructed in the space group algorithm conforms with those described in the International Tables (Hahn (2002)) so that standard cells for comparison with the ITC are generated. This is necessary to determine the standard origin and standard Wyckoff sites of the crystal. The unit cell construction takes advantage of the symmetry properties of the crystal class, together with a set of additional metric rules. The unique choice of lattice basis defined in this way is termed the *conventional crystallographic basis* and the generated cell is called the *conventional crystallographic cell*. When multiple conventions exist, such as in the case of rhombohedral and hexagonal R cells, a single cell is selected. In the aforementioned case, the algorithm defaults to selection of the hexagonal R cell in the obverse setting (see Fig. 2.1). In the case of monoclinic crystals, the default is to select the cell choice 1. In every case, the high-site symmetry origin is selected, resulting in the inversion center being shifted along the body diagonal in certain space groups.

The symmetry properties that are used to define the conventional unit cell comprise the seven holohedries in three dimensions (the point groups of the lattices) and the translation groups. The primitive and Minkowski-reduced lattice is computed before proceeding to compute symmetry relations, due to the fact that mathematically equivalent representations in terms of a superlattice may be given and to reduce the set of points required to reveal the lattice symmetry. Without Minkowski reduction, very skew cells would require search over a very large supercell to find all the lattice symmetries. This procedure may result in the lattice having higher symmetry than the point group of the crystal. The classification in the ITC (Hahn (2002)) is based on so-called Bravais flocks (i.e., Bravais lattice types) that are determined by exam-

ining the point and translation group of the crystal. These are uniquely assigned to Bravais classes (and hence to the seven lattice systems) by finding the smallest possible compatible supergroup from the Bravais classes. To find the conventional cell, the basis vectors are chosen along high-symmetry directions with consideration given to metric properties in some cases (i.e., in the monoclinic and triclinic classes), according to rules defined by Donnay (1943). The rules are summarized by the seven lattice systems. In each case, the conventional form of the associated Bravais lattice(s) are given (Fig. 3.7-3.12).

3.4.1 Cubic

The basis vectors are chosen parallel to three equivalent fourfold axes.

3.4.2 Hexagonal

In addition to the use of high-symmetry directions to determine the lattice basis, hexagonal lattices are constrained by metric conditions: The two axes perpendicular to the sixfold axis must be the shortest lattice vectors that span an angle of 120◦. The direction labeled **c** is chosen parallel to the sixfold axes.

3.4.3 Rhombohedral

In the edition of the ITC used here (Hahn (2002)), rhombohedral lattices are defined in two ways, using so-called hexagonal axes and rhombohedral axes. In the *hexagonal version*, the direction labeled **c** is taken along the threefold axis, and the additional two lattice basis vectors are taken along the equivalent twofold axes. The two non-**c** basis vectors span an angle of 120◦ and are oriented using the obverse setting; that is, lattice points will occur at $(2/3, 1/3, 1/3)$ and $(1/3, 2/3, 2/3)$. In the *rhombohedral version*, a primitive cell is selected with the following condition: The three lattice basis vectors are the shortest symmetrically equivalent (with respect to the threefold axis), non co-planar lattice vectors.

3.4.4 Tetragonal

The tetragonal lattice is determined by conditions analogous to hexagonal lattices. The vector labeled **c** is taken along the fourfold axis. The additional two lattice basis vectors are the shortest lattice vectors along perpendicular twofold axes.

3.4.5 Orthorhombic

Orthorhombic lattices contain three twofold axes. These are parallel to the three lattice basis vectors.

3.4.6 Monoclinic

In the ITC two *settings* are given for monoclinic lattices. In the first, the monoclinic symmetry direction is labeled **b**. The basis vectors labeled **a** and **c** are chosen in the plane perpendicular to **b** and so the following metric condition is satisfied:

$$0 \leqslant -2\boldsymbol{a} \cdot \boldsymbol{c} \leqslant \min(a^2, c^2). \tag{3.1}$$

In words, this means that **a** and **c** are the shortest lattice vectors in their plane and span a non-acute angle. The second setting is analogous. The symmetry direction is now labeled **c**. The vectors **a** and **b** satisfy the same condition given for **a** and **c** above.

Additional consideration may be given to the selection of **a** and **c** (**a** and **b**). Three *cell choices* are given in the ITC. If the basis vectors are chosen so **c** = **e**, **a** = **f** (**e** and **f** refer to the glide vector and projection of the centering vector in the monoclinic plane, respectively) with **b** normal to the plane defined by **c** and **a**, the unit cell corresponds to the first cell choice. However, ignoring such restrictions on the labeling other than the selection of shortest vectors in the plane normal to the symmetry direction will result in one of the listed cell choices; the choice may simply not be the first listed.

3.4.7 Triclinic and the reduced cell

The triclinic lattice is described using a reduced cell (Niggli (1928); Eisenstein (1851); Santoro and Mighell (1970); Hahn (2002)) due to the absence of any relevant symmetry characteristics. The metric conditions related to the determination of the reduced cell are as follows.

Right-handedness

The cell must be chosen to be right-handed: $\boldsymbol{a} \times \boldsymbol{b} = \boldsymbol{c}$.

Type

The *type* of the cell is determined by the sign of

$$T = (\boldsymbol{a} \cdot \boldsymbol{b})(\boldsymbol{b} \cdot \boldsymbol{c})(\boldsymbol{c} \cdot \boldsymbol{a}). \tag{3.2}$$

Positive T corresponds to type I. If $T \leqslant 0$, the cell is type II.

Type-I cell conditions

Main conditions:

$$\boldsymbol{a} \cdot \boldsymbol{b} \leqslant \boldsymbol{c} \cdot \boldsymbol{c}; |\boldsymbol{b} \cdot \boldsymbol{c}| \leqslant \frac{1}{2}\boldsymbol{b} \cdot \boldsymbol{b}; |\boldsymbol{a} \cdot \boldsymbol{c}| \leqslant \frac{1}{2}\boldsymbol{a} \cdot \boldsymbol{a} \tag{3.3}$$

$$\boldsymbol{b} \cdot \boldsymbol{c} > 0 \tag{3.4}$$

Special conditions:

$$\text{If } \boldsymbol{a} \cdot \boldsymbol{a} = \boldsymbol{b} \cdot \boldsymbol{b} \text{ then } \boldsymbol{b} \cdot \boldsymbol{c} \leqslant \boldsymbol{a} \cdot \boldsymbol{c} \tag{3.5}$$

$$\text{If } \boldsymbol{b} \cdot \boldsymbol{b} = \boldsymbol{c} \cdot \boldsymbol{c} \text{ then } \boldsymbol{a} \cdot \boldsymbol{c} \leqslant \boldsymbol{a} \cdot \boldsymbol{b} \tag{3.6}$$

$$\text{If } \boldsymbol{b} \cdot \boldsymbol{c} = \frac{1}{2}\boldsymbol{b} \cdot \boldsymbol{b} \text{ then } \boldsymbol{a} \cdot \boldsymbol{b} \leqslant 2\boldsymbol{a} \cdot \boldsymbol{c} \tag{3.7}$$

$$\text{If } \boldsymbol{a} \cdot \boldsymbol{c} = \frac{1}{2}\boldsymbol{a} \cdot \boldsymbol{a} \text{ then } \boldsymbol{a} \cdot \boldsymbol{b} \leqslant 2\boldsymbol{b} \cdot \boldsymbol{c} \tag{3.8}$$

$$\text{If } \boldsymbol{a} \cdot \boldsymbol{b} = \frac{1}{2}\boldsymbol{a} \cdot \boldsymbol{a} \text{ then } \boldsymbol{a} \cdot \boldsymbol{c} \leqslant 2\boldsymbol{b} \cdot \boldsymbol{c} \tag{3.9}$$

Type-II cell conditions

Main conditions:

$$\boldsymbol{a}\cdot\boldsymbol{b} \leqslant \boldsymbol{c}\cdot\boldsymbol{c}; |\boldsymbol{b}\cdot\boldsymbol{c}| \leqslant \frac{1}{2}\boldsymbol{b}\cdot\boldsymbol{b}; |\boldsymbol{a}\cdot\boldsymbol{c}| \leqslant \frac{1}{2}\boldsymbol{a}\cdot\boldsymbol{a} \tag{3.10}$$

$$(|\boldsymbol{b}\cdot\boldsymbol{c}| + |\boldsymbol{a}\cdot\boldsymbol{c}| + |\boldsymbol{a}\cdot\boldsymbol{b}|) \leqslant \frac{1}{2}(\boldsymbol{a}\cdot\boldsymbol{a} + \boldsymbol{b}\cdot\boldsymbol{b}) \tag{3.11}$$

$$\boldsymbol{b}\cdot\boldsymbol{c} \leqslant 0; \boldsymbol{a}\cdot\boldsymbol{c} \leqslant 0; \boldsymbol{a}\cdot\boldsymbol{b} \leqslant 0 \tag{3.12}$$

Special conditions:

$$\text{If } \boldsymbol{a}\cdot\boldsymbol{a} = \boldsymbol{b}\cdot\boldsymbol{b} \text{ then } |\boldsymbol{b}\cdot\boldsymbol{c}| \leqslant |\boldsymbol{a}\cdot\boldsymbol{c}| \tag{3.13}$$

$$\text{If } \boldsymbol{b}\cdot\boldsymbol{b} = \boldsymbol{c}\cdot\boldsymbol{c} \text{ then } |\boldsymbol{b}\cdot\boldsymbol{c}| \leqslant |\boldsymbol{a}\cdot\boldsymbol{b}| \tag{3.14}$$

$$\text{If } |\boldsymbol{b}\cdot\boldsymbol{c}| = \frac{1}{2}\boldsymbol{b}\cdot\boldsymbol{b} \text{ then } \boldsymbol{a}\cdot\boldsymbol{b} = 0 \tag{3.15}$$

$$\text{If } |\boldsymbol{a}\cdot\boldsymbol{c}| = \frac{1}{2}\boldsymbol{a}\cdot\boldsymbol{a} \text{ then } \boldsymbol{a}\cdot\boldsymbol{b} = 0 \tag{3.16}$$

$$\text{If } |\boldsymbol{a}\cdot\boldsymbol{b}| = \frac{1}{2}\boldsymbol{a}\cdot\boldsymbol{a} \text{ then } \boldsymbol{a}\cdot\boldsymbol{c} = 0 \tag{3.17}$$

$$\text{If } (|\boldsymbol{b}\cdot\boldsymbol{c}| + |\boldsymbol{a}\cdot\boldsymbol{c}| + |\boldsymbol{a}\cdot\boldsymbol{b}| = \frac{1}{2}(\boldsymbol{a}\cdot\boldsymbol{a} + \boldsymbol{b}\cdot\boldsymbol{b}) \text{ then } \boldsymbol{a}\cdot\boldsymbol{a} \leqslant 2|\boldsymbol{a}\cdot\boldsymbol{c}| + \boldsymbol{a}\cdot\boldsymbol{b} \tag{3.18}$$

3.5 The standard crystallographic origin

The convention for origin selection as described in the ITC is based on the choice of high site-symmetry points or inversion centers as origin (Table 3.1). When the site of highest symmetry differs from the inversion center, two entries are displayed in the ITC, differentiated as *origin choice 1* and *origin choice 2*. To be consistent across the space groups, the first origin choice is the origin of choice in the implementation of the routine. Thus, one finds that certain centrosymmetric space groups in the tetragaonal and cubic crystal systems have the inversion center displaced along the body diagonal.

Determination of the standard origin is the final step in the space group search itself: If an origin shift exists such that the input structure symmetry operators become identical to the operators associated with an entry in the ITC, both the space group and the associated origin are found. Origin place a significant role in the representation of symmetry operations. The representation of screw axes, glide planes, rotoinversions, and the other symmetry operations is origin dependent. In order to compare and algebraically solve for the origin shift from one cell to another, the space group symmetries are decomposed into a point operation and a translation.

The standard tables provide this decomposition, not with the explicit display of the symmetry operators (which are origin dependent) but with the Wyckoff general position. The general point (x, y, z) and its equivalent positions can be easily separated into rotational and translational parts. This is done by inspection (automated by the code), pulling off the constant translational shift component by component. If the character of the rotational components of the operations match, full equivalence can be proven or disproven by the existence of a constant shift equating the translational components.

Table 3.1: The standard origins of the 230 space groups with point group and crystal system from ITC (Hahn (2002)). The notation used borrows elements from the Hermann-Mauguin scheme: Rotations, rotoinversions, reflections, and screw rotations are defined equivalently (e.g., for screw rotation, 4_1 denotes a fourfold axis followed by a translation of 1/4 of the lattice vector parallel to the axis.) The symmetry symbols listed for each space group comprise those sufficient for defining the origin. Clusters of symmetry operations (e.g., 222) indicate the intersection of the constituent operations. In some cases more than are necessary for specification of the origin are listed for completeness (e.g., two twofold axes unambiguously define an origin but three may exist). Space groups with (*) are those for which the point group operations are not used alone to determine the standard origin. Boxed point group symbols are the centrosymmetric groups—in these cases, an inversion center is always used as the origin.

Cryst. Syst.	Pnt. Group	Standard Origin (SG number)
Triclinic (1-2)	1	arbitrary**(1)**
	$\boxed{\bar{1}}$	$\bar{1}$**(2)**

Monoclinic (3-15)	2	2**(3)**, 2_1**(4*)**, 2**(5)**
	m	m**(6)**, g**(7*)**, m**(8)**, g**(9*,MB)**
	$\boxed{2/m}$	$\bar{1}$**(10-15)**
Orthorhombic (16-74)	222	222**(16)**, 2_12**(17*)**, $2(2_1 \times 2_1)$**(18*)**, $M(2_1,2_1,2_1)$**(19*)**, 22_1**(20*)**, 222**(21-23)**, $M(2,2,2)$**(24)**
	$mm2$	$mm2$**(25)**, $mg2_1$**(26*)**, $gg2$**(27*)**, $g2$**(28*)**, $g2$**(30*)**, mg**(31*)**, 2**(32)**, 2_1**(33*)**, 2**(34)**, $mm2$**(35)**, $mg2_1$**(36*)**, $gg2$**(37*)**, $mm2$**(38)**, $gg2$**(39*)**, $g2$**(40*)**, $g2$**(41*)**, $mm2$**(42)**, 2**(43)**, $mm2$**(44)**, $gg2$**(45*)**, $gg2$**(46*)**
	$\boxed{mmm}$	$\bar{1}$**(47-74)**
Tetragonal (75-142)	4	4**(75)**, 4_1**(76*)**, 2**(77)**, 4_3**(78*)**, 4**(79)**, 2**(80)**
	$\bar{4}$	$\bar{4}$**(81-82)**
	$\boxed{4/m}$	$\bar{1}$**(83-88)**
	422	422**(89)**, 222**(90)**, 24_1**(91*)**, 22_1**(92*)**, 222**(93)**, 222**(94)**, 24_3**(95*)**, 22_1**(96*)**, 422**(97)**, 222**(98)**
	$4mm$	$4mm$**(99)**, $4(\parallel 2)g$**(100*)**, $2mm$**(101-102)**,$4(\parallel 2)gg$**(103*)**, $4(\parallel 2)g$**(104*)**, $2mm$**(105)**, 2**(106)**, $4mm$**(107)**, $4(\parallel 2)gg$**(108)**, $2mm$**(109)**, $2g$**(110)**
	$\bar{4}2m$	$\bar{4}2m$**(111)**, $\bar{4}g$**(112*-114*)**,$\bar{4}m2$**(115)**, $\bar{4}g$**(116)**, $\bar{4}2_1$**(117*)**, $\bar{4}$**(118)**, $\bar{4}m2$**(119)**, $\bar{4}g2_1$**(120*)**, $\bar{4}2m$**(121)**, $\bar{4}$**(122)**
	$\boxed{4/mmm}$	$\bar{1}$**(123-142)**
Trigonal (143-167)	3	3**(143)**, 3_1**(144*)**, 3_3**(145*)**, 3**(146)**
	$\boxed{\bar{3}}$	$\bar{1}$**(147-148)**
	32	32**(149)**, 32**(150)**, 23_1**(151*-152*)**, 23_2**(153*-154*)**, 32**(155)**
	$3m$	$3m$**(156-157)**, $3g$**(158*-159*)**, $3m$**(160)**, $3g$**(161*)**
	$\boxed{\bar{3}m}$	$\bar{1}$**(162-167)**
Hexagonal (168-194)	6	6**(168)**, 6_1**(169)**, 6_5**(170*)**, 2**(171-172)**, 3**(173)**
	$\bar{6}$	$\bar{6}$**(174)**
	$\boxed{6/m}$	$\bar{1}$**(175-176)**
	622	622**(177)**, 26_1**(178*)**, 26_5**(179*)**, 222**(180-181)**, 32**(182)**
	$6mm$	$6mm$**(183)**, $6gg$**(184*)**, $3m$**(185-186)**

	$\bar{6}m2$	$\bar{6}m2$**(187)**, 32**(188)**, $\bar{6}2m$**(189)**, 32**(190)**,
	$\boxed{6/mmm}$	$\bar{1}$**(191-194)**
Cubic (195-230)	23	23**(195-197)**, $3M(2_1, 2_1, 2_1)$**(198*)**, $3M(2, 2, 2)$**(199)**
	$\boxed{m\bar{3}}$	$\bar{1}$**(200-206)**
	432	432**(207)**, 23**(208)**, 432**(209)**, 23**(210)**, 432**(211)**, $3M(4_3, 4_3, 4_3)$ **(212*)**, $3M(4_1, 4_1, 4_1)$ **(213*-214*)**
	$\bar{4}3m$	$\bar{4}3m$ **(215-217)**, 23 **(218)**, 23 **(219)**, $3M(\bar{4}, \bar{4}, \bar{4})$ **(220)**
	$\boxed{m\bar{3}m}$	$\bar{1}$ **(221-230)**

3.6 Comparison with standard library

An arbitrary crystal structure, once represented in the conventional form (with unknown origin), can be compared to the space group library through application of the relevant point symmetry operations and testing for possible fractional translations. If an isometry exists, it is stored as a symmetry of the crystal for search across the library. The crystal system, point group and Bravais lattice restrict the search to at most sixteen space groups (this occurs in the orthorhombic crystal system with point group mmm) (see Table 3.2). Further restricting the search across the group generators simplifies the routine, although the existence of centering operations can introduce additional complexity. Finally, once the set of candidate space groups with matching crystal system, point group, Bravais lattice, and generator characters are known, a shift of origin is sought algebraically. The existence of a solution will imply a unique space group. With the space group and standard origin known, equivalent sets of atoms (found by application of the full symmetry set) can be used to determine the Wyckoff position parameters and labels.

The crystal point group (crystal class) itself can be uniquely determined using a look-up table given the number of occurrences of the various symmetries (Table

3.3). The characters are associated with the sets of point operator matrices (see column three of Table 3.4) for each crystal system. The groupings of candidate point operations are applied in a manner consistent with the group-subgroup relations of the crystal point groups as described previously.

3.7 Going from lattice symmetry to full crystal symmetry

Computing the symmetry of the lattice is the only completely general process within the space group determination algorithm. *A priori* determination of the full space group would not only be slower, but would add an unnecessary step when the objective is delivering an output consistent with the International Tables. As has been discussed above, the lattice symmetry is not generally consistent with the full point group and space group. The task is then to systematically arrive at the correct representation knowing only the lattice symmetry, which may be incorrect.

The solution lies in the fact that the point group of the lattice *cannot* be lower than the point group of the crystal. Symmetry operations (isometries) of the set of face normals can only be reduced by the existence of interior points. If the contrary is true, the chosen lattice basis was not appropriately reduced to a primitive representation. Thus, be choosing the cell implied by the lattice symmetry as the 0th conventional cell and applying the set of point operations for the lattice on the full crystal (Table 3.7), we can iterate to the correct conventional cell and full set of symmetry operations. The evolution of the conventional cell follows the group-subgroup relations between the crystallographic point groups (Fig. 3.13). For example, cubic may become tetragonal, trigonal, orthorhombic, monoclinic, and triclinic, but tetragonal cannot become cubic, and so forth.

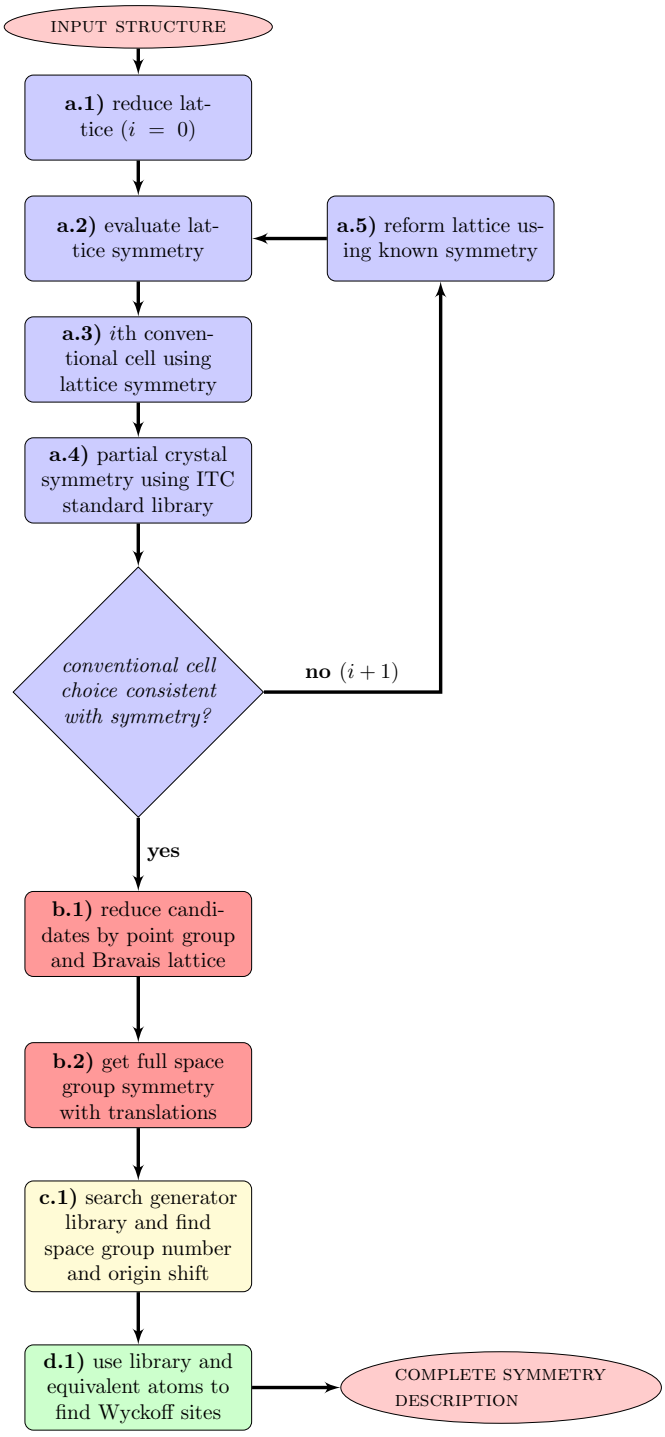


FIGURE 3.1: Flow chart describing the basic design of the space group algorithm

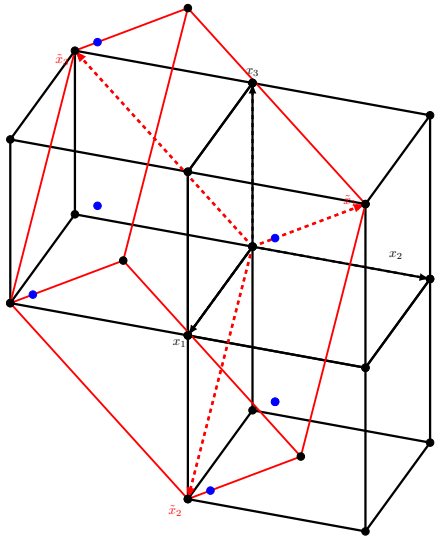


FIGURE 3.2: A cubic lattice with an atom at the origin and an atom displaced along the body diagonal will result in a trigonal crystal. Evaluating the lattice symmetry alone would imply a cubic crystal.

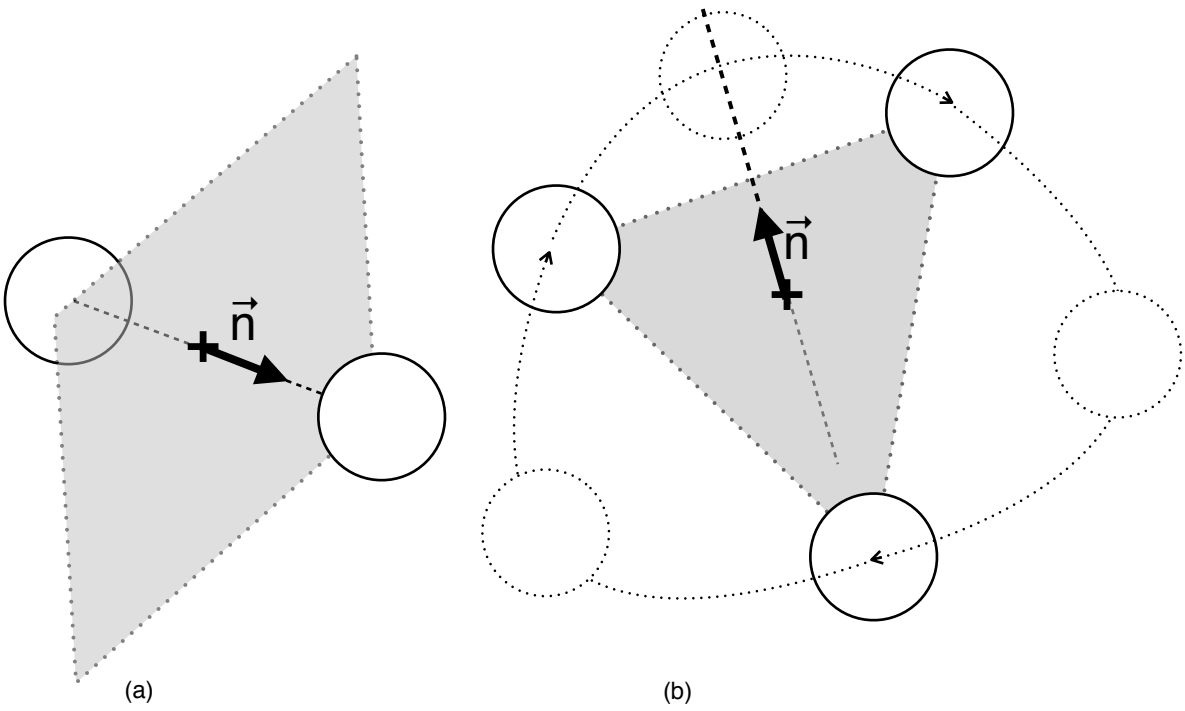


FIGURE 3.3: Illustration of the method used to find point group operations of the lattice. **a)** shows a mirror reflection requiring selection of two atoms. The plane normal is defined by the line between the two points and intersects at the midpoint. **b)** shows a three-fold rotation about an axes. The axes is uniquely determined by the normal to the plane intersecting at the center of mass. This is extended to other crystallographic point operations ($n = 1, 2, 3, 4, 6$) by reduction of co-planer atoms.

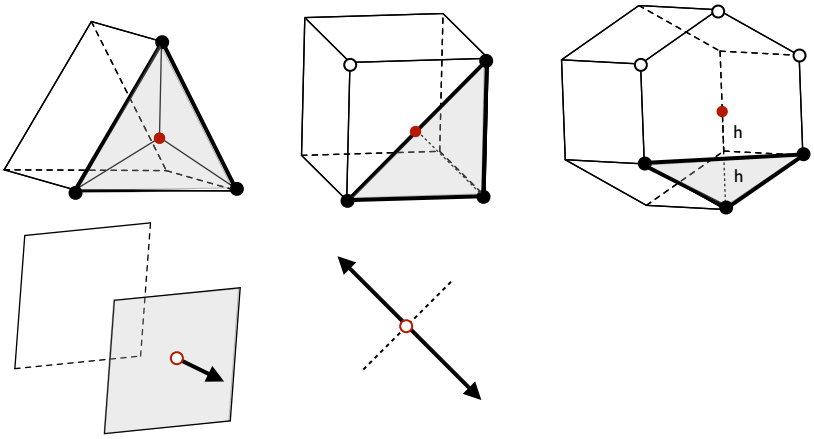


FIGURE 3.4: The geometric objects related to the crystallographic symmetry operations in three dimensions: equilateral triangle, isosceles right triangle, 120-30-30 triangle, plane, and line. Small solid points indicate the intersection of an axis of symmetry with the plane defined by three points.

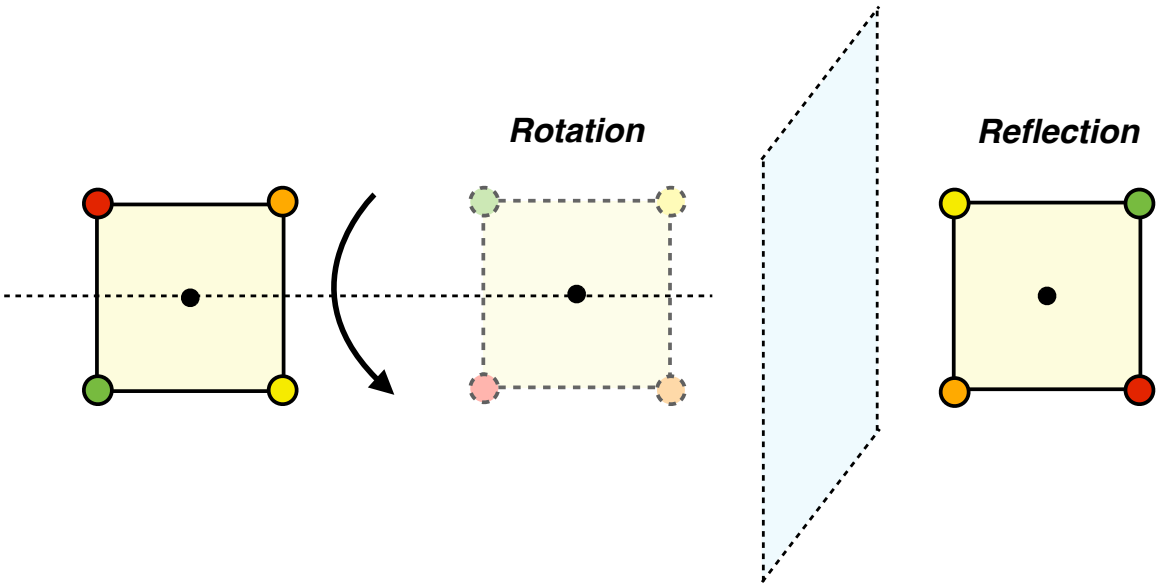


FIGURE 3.5: Cartoon illustrating a rotation-reflection/rotation-inversion operation.

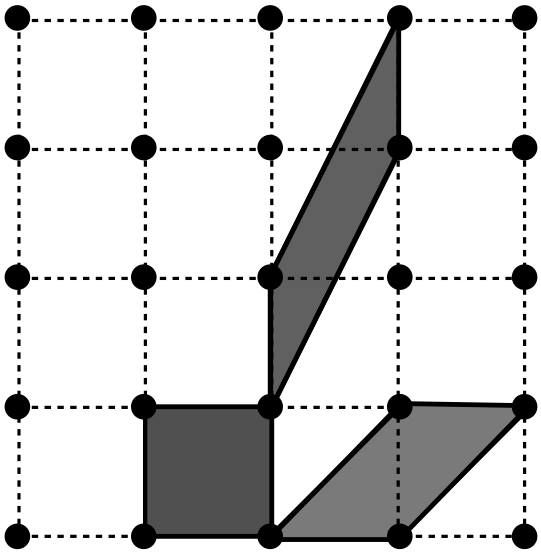


FIGURE 3.6: Three choices of reduced basis for a square lattice. Each has the same, minimal volume.

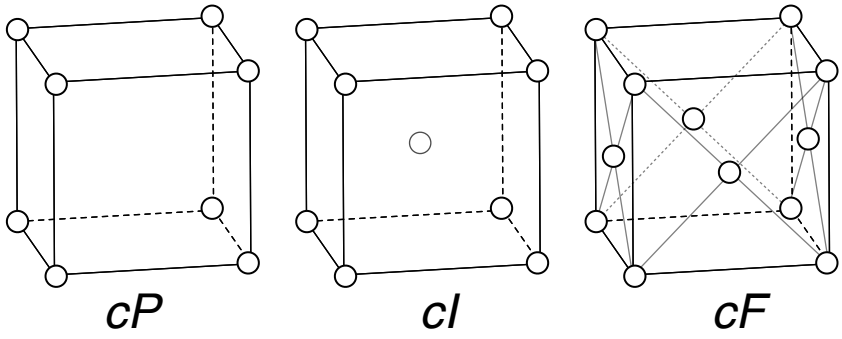


FIGURE 3.7: The Bravais lattices for the cubic lattice system.

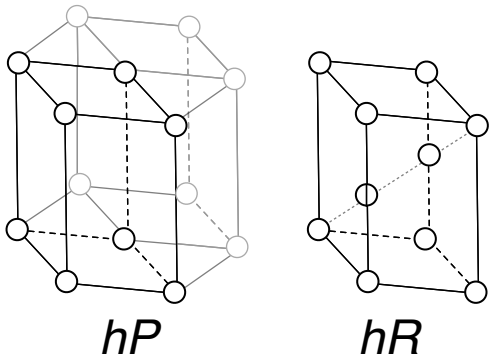


FIGURE 3.8: The Bravais lattices for the hexagonal (hP) and rhombohedral (hR) lattice systems.

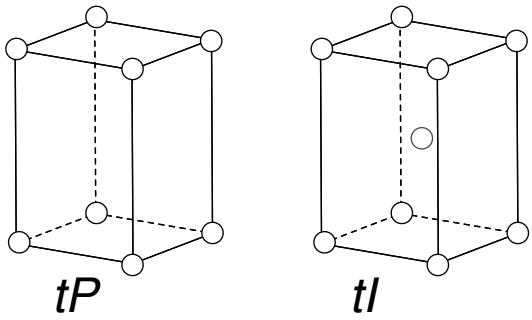


FIGURE 3.9: The Bravais lattices for the tetragonal lattice system.

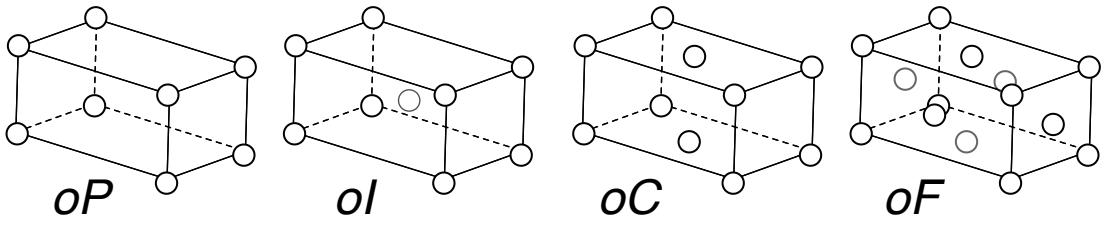


FIGURE 3.10: The Bravais lattices for the orthorhombic lattice system.

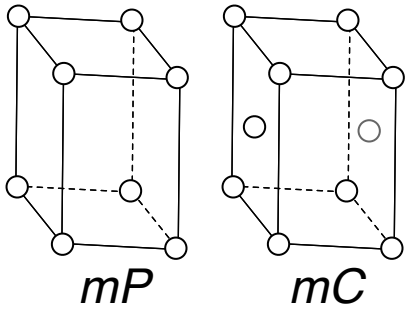


FIGURE 3.11: The Bravais lattices for the monoclinic lattice system.

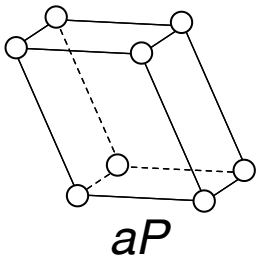


FIGURE 3.12: The Bravais lattice for the triclinic lattice system.

Table 3.2: The crystal point group and Bravais centering allow for a significant reduction of the candidate space group set. The majority of point group and centering pairs limits the space group candidate set to four or fewer. At most the candidate set includes sixteen space groups.

Crystal System	**Point Group**	P	C	F	I
Triclinic	1	1			
	$\bar{1}$	2			
Monoclinic	2	3-4	5		
	m	6-7	8-9		
	$2/m$	10,11,13,14	12,15		
Orthorhombic	222	16-19	20-21	22	23-24
	$mm2$	25-34	35-41	42-43	44-46
	mmm	47-62	63-68	69-70	71-74
Tetragonal	4	75-78			79-80
	$\bar{4}$	81			82
	$4/m$	83-86			87-88
	422	89-96			97-98
	$4mm$	99-106			107-110
	$\bar{4}2m$	111-118		119-122	
	$4/mmm$	123-138			139-142
Trigonal	3	143-145	146		
	$\bar{3}$	147	148		
	32	149-154	155		
	$3m$	156-159	160-161		
	$\bar{3}m$	162-165	166-167		
Hexagonal	6	168-173			
	$\bar{6}$	174			
	$6/m$	175-176			
	622	177-182			
	$6mm$	183-186			
	$\bar{6}m2$	187-190			
	$6/mmm$	191-194			
Cubic	23	195,198		196	197,199
	$m\bar{3}$	200-201, 205		202-203	204,206
	432	207-208,212-213		209-210	211,214
	$\bar{4}3m$	215,218		216,219	217,220
	$m\bar{3}m$	221-224		225-228	229-230

Table 3.3: The characters for a single centered set are enumerated in the point group calculation routine. The identity operation is not included. Thus, for example, the total of the columns in the table below for the highest symmetry cubic point group $m3m$ is 47. (There are a possible 47 unique crystallographic point operations for cubic crystals, not including the identity operator).

Crystal System	**Character Set**										**Point Group**
Triclinic	0	0	0	0	0	0	0	0	0	0	1
	1	0	0	0	0	0	0	0	0	0	$\bar{1}$
Monoclinic	0	0	1	0	0	0	0	0	0	0	2
	0	1	0	0	0	0	0	0	0	0	m
	1	1	1	0	0	0	0	0	0	0	$2/m$
Orthorhombic	0	0	3	0	0	0	0	0	0	0	222
	0	2	1	0	0	0	0	0	0	0	$mm2$
	1	3	3	0	0	0	0	0	0	0	mmm
Tetragonal	0	0	1	0	2	0	0	0	0	0	4
	0	0	1	0	2	0	0	0	2	0	$\bar{4}$
	1	1	1	0	2	0	0	0	2	0	$4/m$
	0	0	5	0	2	0	0	0	0	0	422
	0	4	1	0	2	0	0	0	0	0	$4mm$
	0	2	3	0	0	0	0	0	2	0	$\bar{4}2m$
	1	5	5	0	2	0	0	0	2	0	$4/mmm$
Trigonal	0	0	0	2	0	0	0	0	0	0	3
	1	0	0	2	0	0	0	2	0	0	$\bar{3}$
	0	0	0	3	2	0	0	0	0	0	32
	0	3	0	2	0	0	0	0	0	0	$3m$
	1	3	3	2	0	0	0	2	0	0	$\bar{3}m$
Hexagonal	0	1	0	2	0	2	0	0	0	0	6
	0	1	0	2	0	0	0	0	0	2	$\bar{6}$
	1	1	1	2	0	2	0	2	0	2	$6/m$
	0	0	7	2	0	2	0	0	0	0	622
	0	6	1	2	0	2	0	0	0	0	$6mm$
	0	4	3	2	0	0	0	0	0	2	$\bar{6}m2$
	1	7	7	2	0	2	0	2	0	2	$6/mmm$
Cubic	0	0	3	8	0	0	0	0	0	0	23
	1	3	3	8	0	0	0	8	0	0	$m\bar{3}$
	0	0	9	8	6	0	0	0	0	0	432
	0	6	3	8	0	0	0	0	6	0	$\bar{4}3m$
	1	9	9	8	6	0	0	8	6	0	$m\bar{3}m$
	$\bar{1}$	m	2	3^+ or 3^-	4^+ or 4^-	6^+ or 6^-	$\bar{2}$	$\bar{3}^+$ or $\bar{3}^-$	$\bar{4}^+$ or $\bar{4}^-$	$\bar{6}^+$ or $\bar{6}^-$	

Table 3.4: A transformation to a conventional cell permits a search across a predefined set of operations. The (potentially higher) symmetry of the lattice defines the initial cell choice. If, upon application, the symmetry operations below are not commensurate with the symmetry of the lattice, the cell is reformed to one of lower symmetry. The special case of cubic to trigonal must also be checked.

Crystal System	Essential Symmetry	Point group symmetry operations
Cubic	23	**(1)** 1; **(2)** 3^+ x,x,x; **(3)** 3^- x,x,x; **(4)** $\bar{1}$ $0,0,0$; **(5)** $\bar{3}^+$ x,x,x; **(6)** $\bar{3}^-$ x,x,x; **(7)** 2 $0,0,z$; **(8)** 3^+ $x,\bar{x},\bar{x}$; **(9)** 3^- $x,\bar{x},\bar{x}$; **(10)** 2 $x,x,0$; **(11)** 2 $x,\bar{x},0$; **(12)** 4^+ $0,0,z$; **(13)** 4^- $0,0,z$; **(14)** m $x,y,0$; **(15)** $\bar{3}^+$ $x,\bar{x},\bar{x}$; **(16)** $\bar{3}^-$ $x,\bar{x},\bar{x}$; **(17)** m $x,\bar{x},z$; **(18)** m x,x,z; **(19)** $\bar{4}^+$ $0,0,z$; **(20)** $\bar{4}^-$ $0,0,z$; **(21)** 2 $0,0,z$; **(22)** 3^+ $\bar{x},x,\bar{x}$; **(23)** 3^- $\bar{x},x,\bar{x}$; **(24)** 2 $x,0,x$; **(25)** 2 $\bar{x},0,x$; **(26)** 4^+ $0,y,0$; **(27)** 4^- $0,y,0$; **(28)** m $x,0,z$; **(29)** $\bar{3}^+$ $-x,x,-x$; **(30)** $\bar{3}^-$ $-x,x,-x$; **(31)** m $\bar{x},y,x$; **(32)** m x,y,x; **(33)** $\bar{4}^+$ $0,y,0$; **(34)** $\bar{4}^-$ $0,y,0$; **(35)** 2 $x,0,0$; **(36)** 3^+ $\bar{x},\bar{x},x$; **(37)** 3^- $\bar{x},\bar{x},x$; **(38)** 2 $0,y,y$; **(39)** 2 $0,y,\bar{y}$; **(40)** 4^+ $x,0,0$; **(41)** 4^- $x,0,0$; **(42)** m, $0,y,z$; **(43)** $\bar{3}^+$ $\bar{x},\bar{x},x$; **(44)** $\bar{3}^-$ $\bar{x},\bar{x},x$; **(45)** m $x,y,\bar{y}$; **(46)** m x,y,y; **(47)** $\bar{4}^+$ $x,0,0$; **(48)** $\bar{4}^-$ $x,0,0$
Tetragonal	4 or $\bar{4}$	**(1)**; **(4)**; **(7)**; **(10)**; **(11)-(14)**; **(17)-(21)**; **(28)**; **(35)**; **(42)**
Orthorhombic	222 or (mm)2	**(1)**; **(4)**; **(7)**; **(14)**; **(21)**; **(28)**; **(35)**; **(42)**
Monoclinic	2 or m	**(1)**; **(4)**; **(7)**; **(14)**
Triclinic	none	**(1)**; **(4)**
Hexagonal	6 or $\bar{6}$	**(1)** 1; **(2)** 2 $0,0,z$; **(3)** 2 $x,x,0$; **(4)** 2 $x,\bar{x},0$; **(5)** $\bar{1}$ $0,0,0$; **(6)** m $x,y,0$; **(7)** m $x,\bar{x},z$; **(8)** m x,x,z; **(9)** 3^+ $0,0,z$; **(10)** 6^+ $0,0,z$; **(11)** 2 $x,0,0$; **(12)** 2 $x,2x,0$; **(13)** $\bar{3}^+$ $0,0,z$; **(14)** $\bar{6}^+$ $0,0,z$; **(15)** m $x,2x,z$; **(16)** m $x,0,z$; **(17)** 3^- $0,0,z$; **(18)** 6^- $0,0,z$; **(19)** 2 $0,y,0$; **(20)** 2 $0,y,0$; **(21)** 2 $2x,x,0$; **(22)** $\bar{3}^-$ $0,0,z$; **(23)** $\bar{6}^-$ $0,0,z$; **(24)** m $2x,x,z$; **(25)** m $0,y,z$
Trigonal	3 or $\bar{3}$	**(1)-(9)**; **(11)-(13)**; **(15)-(17)**; **(19)-(21)**; **(24)-(25)**

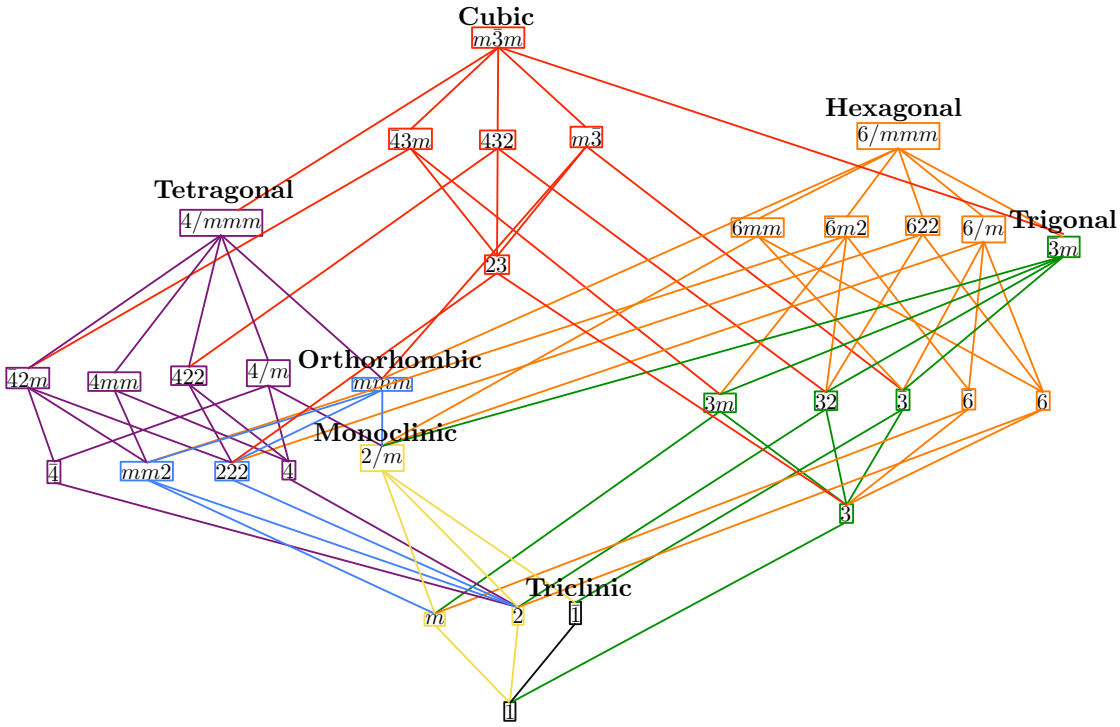


FIGURE 3.13: Group-subgroup relations between the crystallographic point groups and the crystal systems. Point groups belonging to the same crystal system are represented by the same color.

4

Functional descriptions of symmetry routines in the space group determination algorithm

In this chapter I provide a detailed description of the various functions and classes written in C++ that comprise the symmetry routines. The functions called to determine the symmetry of an arbitrary input structure are described in the order in which they are utilized. Prior to this functional description, the underlying class structures are described. The routines are built around two primary classes, `class CrystalStructure`, which contains functions for manipulation, conversion, expansion, etc. of the crystal structure and the `class CrystalSymmetry` for handling the symmetry related information. Supporting data structures are also used: `class _atom`, `class wyckoffsite`, `class xmatrix`, `class xvector`, `class xstructure`, and the symmetry classes for glides, screws, rotoinversions, and translations described in Chapter 3. The `xstructure` class is the underlying data structure of the AFLOW software and contains as a small subset many of the `class CrystalStructure` member functions. Due to the size of the `xstructure` class, and the requirement of frequent passing between functions in the symmetry rou-

tines as well as different needs for the overloaded operators, the smaller and narrower `CrystalStructure` class was designed. For purposes of full integration into the high-throughput framework, the higher-level functions (the main space group, Wyckoff site, point group, crystal class, and symmetry-set functions) have associated member functions and variables in `xstructure`. The two vector classes `xvector` and `xmatrix` are the native vector classes used in AFLOW with the typical functionality (matrix addition and multiplication, inversion, row-reduction etc).

4.1 Some important classes used in symmetry routines

4.1.1 Crystal structure

The `CrystalStructure` class comprises member functions and variables for storing and manipulating input structure-related data. The principal member functions are `void get_lattice(vector<string> [vs])` and `void reduce_lattice()`. The `vector<string> [vs]` argument of the `get_lattice` function is the standard POSCAR file format broken by line into the C++ `std::string` class. The `get_lattice` function fills the `CrystalStructure` member variables for handling the structure-related data: e.g., `vector<_atom> atomic_basis_`, `xmatrix<double> lattice`, etc. Additional member functions and variables include those that expand the lattice and atomic basis by specified multipliers, switch the atomic basis vectors from Cartesian to direct coordinates, and store critical information (lattice system, crystal system, centering operations, chemical labels, etc.).

4.1.2 Crystal symmetry

The calculated symmetry information is stored in the `CrystalSymmetry` class. This is in turn passed to an instance of the main structure class of AFLOW. Currently, the public members of `CrystalSymmetry` also include structural information, namely, the lattice, title, and volume. The symmetry related information that is stored includes the Wyckoff site parameters and symbols, the space group, point group, Bravais

lattice, crystal system, and point group. If necessary, the class can be extended to include the explicit representation of the symmetry operations computed in the space group routine. This data is computed and displayed but not stored in the class.

4.1.3 Atoms in crystals

The `_atom` class is a lightweight and simple data structure designed to easily handle the information related to defining an atom inside a crystal, mainly coordinate and chemical type. By using overloaded operators and member functions, stream insertion '<<', frequent comparisons in loops for symmetry searches and the process of coding and debugging, were greatly simplified.

4.1.4 Wyckoff sites

The `wyckoffsite` class is similar to the `_atom` class described above but includes a member variable, a `std::string` to store the ITC standard Wyckoff symbol. Although both `_atom` and `wyckoffsite` contain

`xvector<double> coordinate`, it must be understood that the coordinate associated with a Wyckoff site describes, in general, a *set* of atomic positions whereas the coordinate in `_atom` describes the location of a single atom.

4.2 Algorithm Functions

4.2.1 Conventional Cell: Lattice Reduction and Lattice Symmetry

Minkowski Lattice Reduction

Upon entering the main loop for the determination of the standard space group symbol and associated information, the first matter is the primitivization of the lattice and extraction of structure data into the code-readable format. Calculation of the standard conventional cell follows execution of these functions. Functions for the extraction of lattice information from the input structure file, reduction of the lattice, and computation of the lattice symmetry also comprise the first part of the conventional cell routine, which is iterated in a way that has been described in earlier

chapters.

For the reduction routine, one might expect to naively expand the crystal structure (both the lattice and the atomic basis) and search across the set of translations defined by the differences between atom sites. Those translations that preserve the original structure may be stored as candidate lattice vectors. Then, from the candidate set, combinations of three linearly independent vectors with minimum determinant will define a primitive cell.

Although this does return a primitive cell and will often be sufficient for the algorithm discussed here, as the input structure's lattice basis is chosen to be more skew the expansion over which the lattice symmetry search is conducted must increase accordingly. Figure 4.1 illustrates this effect in a simple 2-D case for a square lattice. Practically, this can cause the routine to be prohibitively slow for very skew cells and, importantly, this can be avoided by additional constraints on the choice of primitive cell.

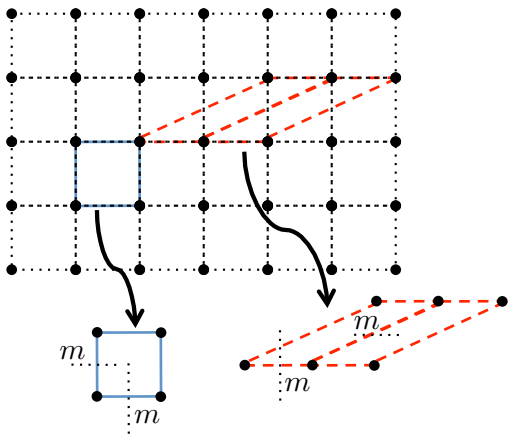


FIGURE 4.1: Two choices for the primitive cell of a cubic lattice. The non-orthogonal cell must be doubled to reveal the equivalent symmetry information of the orthogonal cell. Here, only the mirror/2-fold symmetry operations are shown, although the same effect can occur for any symmetry operation.

This problem is solved by including an additional transformation of the lattice

representation. Specifically, the primitive lattice is transformed using a Minkowski reduction algorithm (Buchmann (2004))—the choice of lattice vectors that have the shortest possible magnitude. The Minkowski basis is computed using a greedy algorithm which selects the locally optimal choice within the given constraints. The objective function of the greedy reduction is the orthogonality defect, for which the Minkowski reduced lattice basis will have a minimum. The orthogonality defect is simply the ratio of the product of the magnitudes of the basis vectors (volume of completely orthogonal cell) to the determinant of the cell (volume of the true cell), and hence, $\delta(L) \geqslant 1$. As the lattice basis is chosen so that the vectors are "more" orthogonal the denominator will approach the numerator. Equality with unity is reached only when the basis is orthogonal.

$$\delta(L) = \frac{\prod ||l_i||}{\det(L)} \tag{4.1}$$

Lattice Symmetry

Once the lattice is transformed to the Minkowski representation, the lattice symmetry is computed directly without reference to any particular orientation. This is done using the topological method described in earlier chapters. Three functions are used: one for mirror operations, triplet operations, and two-fold operations. As described previously and illustrated in Figure 3.4, the lattice symmetries can be found by searching sets of two and three lattice points across a finite subset of the lattice. The set of equivalent points nearest in proximity (spanning the lattice) defines a reasonable set of points that reveals the lattice symmetry with minimal expansion. Having performed Minkowski reduction, an expansion (or supercell) along each direction by multiplicity one is always sufficient to find the symmetry of the lattice. Although this may not strictly be the most efficient method to obtain the lattice symmetry, the simplicity of the algorithm together with the minimal upfront overhead (the primitive cell of a Bravais lattice always only contains one point, so

even a large expansion is tractable) make it desirable within this context.

Using the three functions for lattice symmetry, the crystal structure is initially categorized into one of the seven lattice systems (cubic, hexagonal, rhombohedral, tetragonal, orthorhombic, monoclinic, and tetragonal) using the mirror operations alone. This is done early in the algorithm to restrict types for the construction of the conventional cell: cubic is identified by nine mirror operations ($m\bar{3}m$), hexagonal by seven mirror operations($6/mmm$), rhombohedral by three mirror operations ($\bar{3}m$), tetragonal by five mirror operations ($4/mmm$), orthorhombic by three mutually orthogonal mirror operations (mmm), monoclinic by one mirror operation ($2/m$), and triclinic by the existence of none ($\bar{1}$).

4.2.2 Conventional Cell

Once the symmetry of the underlying lattice is known, the first iteration of the standard conventional cell is computed according to the rules outlined in Chapter 3. The cubic-lattice structures (identified by nine mirror operations) are constructed by identifying the four-fold axes. These are found by searching the `vector<Screw>` return of the `rot_ops_vec` function, a vector of `Screw` class objects, for rotations of order four. The lattice vector magnitudes are determined by using the `closest_point` function. This computes the closest lattice point along a given direction vector $\vec{d}$ and lattice $\mathbf{L}$.

The closest-point routine is essentially one which solves the greatest common divisor problem, because the direction vector expressed in terms of the lattice basis can be equivalently represented with arbitrary accuracy by an integer vector $\vec{d_I}$. One can then solve the problem of finding the greatest common divisor for the components of the integer approximation of the direction vector.

For example, consider a cubic lattice with unit vectors and a direction vector (in Cartesian coordinates) of $(0.10, 0.15, 0.60)$. Because the lattice is cubic with

volume one, the direct representation of the direction vector is also $(0.10, 0.15, 0.60)$, and the integer representation is $(10, 15, 60)$. The greatest common divisor of these components is 5. Thus, the closest lattice point with reference to a lattice point origin is at $(2, 3, 12)$ with respect to the cubic lattice. The direction in Cartesian coordinates can be found by multiplying by the lattice basis matrix—multiplying a vector in the Cartesian basis by the lattice matrix inverse gives the vector in "direct" coordinates.

Using the newly constructed conventional cell, the Bravais lattice and the conventional atomic basis are found. The Bravias lattice is defined by centering operations within the cell, which simply implies the existence of additional *lattice points* within the conventional cell. Thus, the expanded lattice is searched for points lying within the conventional cell, the quantity and coordinate relationships of which will serve to distinguish between the centering types (ie, face-centered, body-centered, etc.).

The conventional atomic basis is found in the same way. The full expanded crystal (using the Minkowski primitive reduction discussed earlier) is searched for points within the conventional cell. These are added to a vector of the `_atom` class containing the points of the conventional basis.

The remaining six lattice symmetry classes are used to construct the first iteration of the conventional cell using similar procedures as those for cubic but following the appropriate prescription for conventional cells outlined in Chapter 3. One point of note for monoclinic and orthorhombic cells: additional checks and manipulations may be necessary to select a C-centered cell (with centered atom at $(1/2, 1/2, 0)$). The algorithm selects the 'b' axis as the monoclinic axis. Also, when lattice directions are defined by equivalent symmetries, ambiguity exists in the definition of the cell. Thus, an additional loop for iterating over these possibilities must also be performed.

4.3 Iterating the Conventional Cell

Because the addition of interior points can only *reduce* or leave unchanged the point group of a crystal, the point group of the lattice does not uniquely determine the point group of the crystal. Thus, the crystal class implied by the lattice may be incorrect. It is entirely possible that the conventional cell constructed in the way described above is not consistent with the International Tables. More regarding this has been said in Chapter 3. The restrictions imposed by the lattice symmetry on the full symmetry are described visually in the group-subgroup diagram (Fig. 3.13). In general, one can only infer that the crystal point group is a *subgroup* of the lattice point group.

In order to get the proper conventional cell, an iterative procedure is employed. The `symfolder check_ccell(CrystalStructure C)` function performs the main consistency check and reforms the stored symmetry operations to reflect the full symmetry of the crystal. The `symfolder` object is a C structure aggregating vectors of symmetry operations (two-fold, n-fold ($n > 2$), and mirror operations) a Boolean variable for passing consistency or inconsistency between functions, and string objects for describing the crystal system and lattice system. The input `CrystalStructure` is the class object from the conventional cell routine containing the lattice vectors and atomic positions as well as other relevant information associated with the conventional cell.

The `check_ccell` function begins by calling external functions for loading vectors containing the point symmetry operations with reference to the standard conventional cells in the ITC. Vectors of indices for describing the *possible* operations as reduced by the lattice type are also loaded. In this way, the full search for symmetry operations of the crystal is greatly reduced by a) conventionalizing the cell and b) knowing the symmetry of the lattice. Extending a library of point operations to the

compound operations that generally exist for crystals is done by considering possible translational shifts within the unit cell. These are simply the differences between atomic positions of equivalent chemical types (a crystal cannot have a symmetry operation that maps an atom of one type to an atom of another type). More specifically, the possible internal translations are defined by the set of difference vectors with reference to any one atomic position. This is because a compound operation (a general symmetry operation) acts equally on all sites of the crystal. If, for example, one has a crystal with N_i atoms of type A_i and only 1 atom of type B the space group must be one of the 73 simple space groups, because there can be no internal translation from B to A_i.

The above fact has been utilized to improve the efficiency of the space group symmetry element search and the lattice primitivization routine: Rather than searching over all site-vector differences, or over differences defined by an arbitrary set of positions for a given chemical type, the search is only performed over the differences defined by the set of equivalent chemical types with smallest cardinality. Furthermore, one need only consider the difference between a particular site i and all others within the set (symmetry operations act on all sites). The increase in efficiency made by these modifications can be significant for larger cells.

The full crystal symmetry is determined within `check_ccell` and thus the crystal class can be computed and compared with the assumed crystal class (from the lattice symmetry). In the event a difference is found, the cell structure is modified to conform with the standard conventions. This is done by passing reloaded vectors of symmetry operations to the conventional cell routine with a label defining the appropriate crystal class. Because the full symmetry is generated within `check_ccell` the routine will require at most one iteration to arrive at the correct conventional cell. The crystal system and crystal point group are determined using the *character set* as described in Table 3.3.

4.4 Symmetry Elements With Respect to Standard Conventional Cell

In the main function of the space group algorithm, `xstructure SpaceGroup( istream & cin )`, the point group and Bravais lattice are used to reduce the set of candidate space groups to at most sixteen possibilities (Table 3.2). An additional reduction of the candidate set is made by referring to a library containing the generator operators of the space groups. The generator matching is defined as equivalence of the point group symbol (e.g., the glides (n,d,etc.) are matched to a mirror plus a shift because of the origin dependence of the glide symbols) and the characteristic symmetry direction.

For space groups with centered lattices (C,I,F) there are centered groups of equivalent symmetry operations within conventional cells. Because the generator operations are selected from a particular centered group, and the groupings depend on the origin choice, the algorithm searches over combinations of possible generator sets to determine a space group match. The actual matching function simply solves a linear problem for an origin shift between the library generator set and the input crystal set. To derive the linear problem that is solved for the origin shift, consider two symmetrically-equivalent points, a and a'. By the appropriate origin shift, $\vec{\mathcal{O}}$, one can transform between the ITC representation and the original coordinate system.

$$\begin{aligned} \vec{a}_{\rm ITC} &= \vec{a}_{\rm P} + \vec{\mathcal{O}} \\ \vec{a}'_{\rm ITC} &= \vec{a}'_{\rm P} + \vec{\mathcal{O}} \end{aligned}$$

Let the purely point-operation component of the symmetry relation transforming a to a' be designated as $\mathbf{R}$. Then, applying $\mathbf{R}$ to the first relation above and taking the difference gives

$$\vec{a}'_{\rm ITC} - \mathbf{R}\vec{a}_{\rm ITC} = \vec{a}'_{\rm P} + \vec{\mathcal{O}} - \mathbf{R}\vec{a}_{\rm P} - \mathbf{R}\vec{\mathcal{O}} \tag{4.2}$$

The difference between the point-transformed site and the fully transformed site is

the internal shift, $\vec{s}$.

$$\vec{s}_{\mathrm{ITC}} = \vec{s}_{\mathrm{P}} + \vec{\mathcal{O}} - \mathbf{R}\vec{\mathcal{O}} \tag{4.3}$$

Simplification leads to the linear matrix problem,

$$(\mathbf{1} - R)\vec{\mathcal{O}} = (\vec{s}_{\mathrm{ITC}} - \vec{s}_{P}) \tag{4.4}$$

The function that is used to solve the linear problem (which is usually over-determined) returns a Boolean variable with value true if a solution exists and writes the origin shift to a vector passed as an argument.

4.5 Standard Origin and Wyckoff Positions

Once the standard origin and space group are known, direct comparison with the library of Wyckoff positions gives the Wyckoff site names and variables. The atomic positions are first grouped into "equivalent sets"—that is, groups of atoms that are related by the symmetry of the space group. These are essentially the Wyckoff positions without the standard labels and parameters. Starting with one atom position, the symmetry operations of the group are applied. When a unique position is generated, it is stored with the original position. This is done for all operations and atoms, until all are grouped.

Similar to solving for the origin shift, finding the Wyckoff parameters involves a linear problem. This is most easily shown by example. Consider the position $(1/4, y, -y + 1/2)$ from the ITC library and an arbitrary position (a, b, c) from the associated symmetry group for the input structure. The Wyckoff parameter "problem" is represented by a set of linear relations requiring equivalence between the

library position and the input structure position with three parameters (x,y,z).

$$\begin{aligned} 0x + 0y + 0z + 1/4 &= a \\ 0x + 1y + 0z + 0 &= b \\ 0x - 1y + 0z + 1/2 &= c \end{aligned} \tag{4.5}$$

(4.6)

A row-reduction algorithm is then applied to the matrix,

$$\left(\begin{array}{ccc|c} 0 & 0 & 0 & a - 1/4 \\ 0 & 1 & 0 & b \\ 0 & -1 & 0 & c - 1/2 \end{array}\right).$$

If, for example $(a, b, c) = (1/4, 0, 1/2)$, the row reduced matrix is,

$$\left(\begin{array}{ccc|c} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right),$$

and the Wyckoff parameter $y = 0$. Similarly, if no solution exists, the row reduced form makes this clear. Such is the case when $(a, b, c) = (1/4, 1/4, 1/2)$ and the row reduced matrix becomes,

$$\left(\begin{array}{ccc|c} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right).$$

A set of conditional logic statements evaluated on the row reduced matrix returns the existence or non existence of a solution: If any of the rows have all zero values for columns one through three and a nonzero value for column four, a solution does not exist because a sum of zeros cannot be nonzero.

4.6 Example applications of the code

Several examples are given in this section to demonstrate the output and functionality of the symmetry code. The structure files used in the examples are taken from the AFLOW structure library (Curtarolo et al. (2012b)).

4.6.1 Example 1-TRIGONAL

This structure (AFLOW prototype number 10) belongs to the trigonal crystal system, has point group $\bar{3}m$, and space group #164-$P\bar{3}m1$:

```
AB/10 - (10) - FCC [A2B] (10)
-246906.115800
  -0.50000000000000   0.00000000000000   0.50000000000000
   0.50000000000000   0.50000000000000   0.00000000000000
  -1.00000000000000   1.00000000000000  -1.00000000000000
2 1
Direct(3) [A2B1]
   0.33333333333333   0.66666666666667   0.66666666666667  A
   0.66666666666667   0.33333333333333   0.33333333333333  A
   0.00000000000000   0.00000000000000   0.00000000000000  B
```

Using the default tolerance for all routines, Minkowski transformation and reducing the lattice leaves the bases unchanged.

```
WARNING: using default tolerance 0.001
Using tolerance: 0.001
MINKOWSKI-REDUCED PRIMITIVE CELL
AB/10 - (10) - FCC [A2B] (10)
-246906.115800
   0.00000000000000   0.50000000000000   0.50000000000000
   0.50000000000000   0.50000000000000   0.00000000000000
  -1.00000000000000   1.00000000000000  -1.00000000000000
2 1
Direct(3) [A2B1]
```

```
0.33333333333333   0.33333333333334   0.66666666666667   A
0.66666666666667   0.66666666666666   0.33333333333333   A
0.00000000000000   0.00000000000000   0.00000000000000   B
```

Analysis of the lattice reveals seven mirror operations; thus, the crystal is conformed to hexagonal/trigonal symmetry. A complete search for symmetry elements using the symmetry library gives the following:

```
latticetypechar: h
symlist: "1 3 3 2 0 0 0 2 0 0 "
Original 'crystal system' / lattice system: HEXAGONAL
CRYSTAL SYSTEM: TRIGONAL
POINT GROUP: -3m
```

The "symlist" provides the number of symmetry elements of type $\bar{1}$, m, 2, $3^{\pm}$, $4^{\pm}$, $6^{\pm}$, $\bar{3}^{\pm}$, $\bar{4}^{\pm}$, and $\bar{6}^{\pm}$, respectively, from which the point group and crystal system are derived (see Table 3.3). Using the above symmetry information, the standard conventional cell is constructed:

```
CONVENTIONAL CELL CHOICE:
AB/10 - (10) - FCC [A2B] (10)
-246906
0.61237243569579 0.35355339059327 0.0000000000000
-0.61237243569579 0.35355339059327 0.0000000000000
0.0000000000000 0.0000000000000 1.7320508075689
2 1
Direct(3)
0.66666666666667 0.33333333333333 0.66666666666667 0
0.33333333333333 0.66666666666667 0.33333333333333 0
0.0000000000000 0.0000000000000 0.0000000000000 1
```

The full symmetry set is known:

```
1          0.0000e+00  0.0000e+00  0.0000e+00
2 x,x,0    0.0000e+00  0.0000e+00  0.0000e+00
-1 0,0,0    0.0000e+00  0.0000e+00  0.0000e+00
m x,-x,z    0.0000e+00  0.0000e+00  0.0000e+00
3+ 0,0,z    0.0000e+00  0.0000e+00  0.0000e+00
2 x,0,0    0.0000e+00  0.0000e+00  0.0000e+00
-3+ 0,0,z    0.0000e+00  0.0000e+00  0.0000e+00
m x,2x,z    0.0000e+00  0.0000e+00  0.0000e+00
3- 0,0,z    0.0000e+00  0.0000e+00  0.0000e+00
2 0,y,0    0.0000e+00  0.0000e+00  0.0000e+00
-3- 0,0,z    0.0000e+00  0.0000e+00  0.0000e+00
m 2x,x,z    0.0000e+00  0.0000e+00  0.0000e+00
```

Given the Bravais type of primitive (P) and the point group ($\bar{3}m$) the candidate space group set is reduced to one of four: #162-#165. Searching the generator sets for a match leaves only space group #164.

The set of standard Wyckoff positions (from the International Tables) is loaded for space group #164:

```
Space Group 164
 (0,0,0)
 12 j 1 (x,y,z) (-y,x-y,z) (-x+y,-x,z) (y,x,-z) (x-y,-y,-z) (-x,-x+y,-z)
(-x,-y,-z) (y,-x+y,-z) (x-y,x,-z) (-y,-x,z) (-x+y,y,z) (x,x-y,z)
 6 i .m. (x,-x,z) (x,2x,z) (-2x,-x,z) (-x,x,-z) (2x,x,-z) (-x,-2x,-z)
 6 h .2. (x,0,.5) (0,x,.5) (-x,-x,.5) (-x,0,.5) (0,-x,.5) (x,x,.5)
 6 g .2. (x,0,0) (0,x,0) (-x,-x,0) (-x,0,0) (0,-x,0) (x,x,0)
 3 f .2/m. (.5,0,.5) (0,.5,.5) (.5,.5,.5)
```

```
 3 e .2/m. (.5,0,0) (0,.5,0) (.5,.5,0)
 2 d 3m. (1/3,2/3,z) (2/3,1/3,-z)
 2 c 3m. (0,0,z) (0,0,-z)
 1 b -3m. (0,0,.5)
 1 a -3m. (0,0,0)
```

Separating the atomic basis into equivalent atom groupings and then solving for the Wyckoff variables gives two positions: 'a' and 'd'. The Wyckoff variables (as a coordinate triple (x,y,z), atom types, labels and site symmetry are given.

```
0.0000e+00  0.0000e+00  3.3333e-01  A  2 d 3m.
0.0000e+00  0.0000e+00  0.0000e+00  B  1 a -3m.
```

Finally, a structure representation using the the conventional cell and the Wyckoff positions (the "WYCCAR") can be displayed (space group number, point group, and Bravais label are included in the title):

```
[TITLE] | SG #164 PG: -3m BL: P
-246906
0.61237243569579 0.35355339059327 0.0000000000000
-0.61237243569579 0.35355339059327 0.0000000000000
0.0000000000000 0.0000000000000 1.7320508075689
2 1
DIRECT(WYCCAR)
 3.3333e-01  6.6667e-01  3.3333e-01 A 2 d 3m.
 0.0000e+00  0.0000e+00  0.0000e+00 B 1 a -3m.
```

4.6.2 Example 2-MONOCLINIC

This structure (AFLOW prototype number 73) belongs to the monoclinic crystal system, has point group $2/m$, and space group #10-$P2/m$:

```
AB/73 - (73) - BCC [AB3] (73)  (htqc library)
-123438.347400
   0.00000000000000  -2.00000000000000   0.00000000000000
  -1.00000000000000   0.00000000000000  -1.00000000000000
   0.50000000000000   2.50000000000000  -0.50000000000000
1 3
Direct(4) [A1B3]
   0.00000000000000   0.00000000000000   0.00000000000000  A
   0.75000000000000   0.50000000000000   0.00000000000000  B
   0.50000000000000   0.00000000000000   0.00000000000000  B
   0.25000000000000   0.50000000000000   0.00000000000000  B
```

Using the default tolerance for all routines, Minkowski transformation and reducing the lattice modifies the basis: (1) $a' \rightarrow a + c$; (2) $c' \rightarrow c - 2a'$

```
WARNING: using default tolerance 0.001
Using tolerance: 0.001
MINKOWSKI-REDUCED PRIMITIVE CELL
AB/73 - (73) - BCC [AB3] (73)  (htqc library)
-123438.347400
   0.50000000000000   0.50000000000000  -0.50000000000000
  -1.00000000000000   0.00000000000000  -1.00000000000000
  -0.50000000000000   1.50000000000000   0.50000000000000
1 3
Direct(4) [A1B3]
   0.00000000000000   0.00000000000000   0.00000000000000  A
   0.25000000000000   0.50000000000000   0.25000000000000  B
   0.50000000000000   0.00000000000000   0.50000000000000  B
```

```
0.75000000000000   0.50000000000000   0.75000000000000  B
```

The orthogonality defect of the transformed basis can be compared with the original: The transformed representation has $\delta = 1.02$ and the original representation has $\delta = 3.67$. The volume remains the same because the original representation is primitive.

Analysis of the lattice reveals one mirror operation; thus, the crystal is conformed to monoclinic symmetry. A complete search for symmetry elements using the symmetry library gives the following:

```
latticetypechar: m
symlist: "1 1 1 0 0 0 0 0 0 0 "
Original 'crystal system' / lattice system: MONOCLINIC
CRYSTAL SYSTEM: MONOCLINIC
POINT GROUP: 2/m
```

Using the known symmetry properties, the standard conventional cell is constructed:

```
CONVENTIONAL CELL CHOICE:
AB/73 - (73) - BCC [AB3] (73)
-123438
-0.70710678118655 0.50000000000000 0.0000000000000
0.0000000000000 0.0000000000000 1.4142135623731
-0.70710678118655 -1.5000000000000 0.0000000000000
3 1
Direct(4)
0.25000000000000 0.50000000000000 0.75000000000000 1
0.50000000000000 0.0000000000000 0.50000000000000 1
0.75000000000000 0.50000000000000 0.25000000000000 1
0.0000000000000 0.0000000000000 0.0000000000000 0
```

The full symmetry set is known:

```
1          0.0000e+00  0.0000e+00  0.0000e+00
-1 0,0,0   0.0000e+00  0.0000e+00  0.0000e+00
2 0,y,0    0.0000e+00  0.0000e+00  0.0000e+00
m x,0,z    0.0000e+00  0.0000e+00  0.0000e+00
```

Given the Bravais type of primitive (P) and the point group $(2/m)$ the candidate space group set is reduced to one of four: #10-#14. Searching the generator sets for a match leaves only space group #10.

The set of standard Wyckoff positions (from the International Tables) is loaded for space group #10:

```
Space Group 10
 (0,0,0)
 4 o 1 (x,y,z) (-x,y,-z) (-x,-y,-z) (x,-y,z)
 2 n m (x,.5,z) (-x,.5,-z)
 2 m m (x,0,z) (-x,0,-z)
 2 l 2 (.5,y,.5) (.5,-y,.5)
 2 k 2 (0,y,.5) (0,-y,.5)
 2 j 2 (.5,y,0) (.5,-y,0)
 2 i 2 (0,y,0) (0,-y,0)
 1 h 2/m (.5,.5,.5)
 1 g 2/m (.5,0,.5)
 1 f 2/m (0,.5,.5)
 1 e 2/m (.5,.5,0)
 1 d 2/m (.5,0,0)
 1 c 2/m (0,0,.5)
 1 b 2/m (0,.5,0)
```

```
 1 a 2/m (0,0,0)
```

Separating the atomic basis into equivalent atom groupings and then solving for the Wyckoff variables gives three positions: 'a', 'g', and 'n'. The Wyckoff variables as a triple (x,y,z), labels, and site symmetry are then given.

```
2.5000e-01  0.0000e+00  7.5000e-01  B  2 n m
0.0000e+00  0.0000e+00  0.0000e+00  B  1 g 2/m
0.0000e+00  0.0000e+00  0.0000e+00  A  1 a 2/m
```

Finally, a structure representation using the the conventional cell and the Wyckoff positions (the "WYCCAR") can be displayed (space group number, point group, and Bravais label are included in the title):

```
[TITLE] | SG #10 PG: 2/m BL: P
-123438
-7.0711e-01  5.0000e-01 -5.5511e-17
-1.1102e-16  0.0000e+00  1.4142e+00
-7.0711e-01 -1.5000e+00 -5.5511e-17
3 1
DIRECT(WYCCAR)
 2.5000e-01  5.0000e-01  7.5000e-01 B 2 n m
 5.0000e-01  0.0000e+00  5.0000e-01 B 1 g 2/m
 0.0000e+00  0.0000e+00  0.0000e+00 A 1 a 2/m
```

4.6.3 Example 3-TETRAGONAL

This structure (AFLOW prototype number 285) belongs to the tetragonal crystal system, has point group $4/m$, and space group #87-$I4/m$:

```
AB/285 - (285) - ???? [AB] (285)  (htqc library)
-123432.463200
```

```
   5.72000000000000   0.00000000000000   0.00000000000000
   0.00000000000000   5.72000000000000   0.00000000000000
   2.86000000000000   2.86000000000000   1.78200000000000
1 4
Direct(5) [A1B4]
   0.00000000000000   0.00000000000000   0.00000000000000  A
   0.20000000000000   0.40000000000000   0.00000000000000  B
   0.60000000000000   0.20000000000000   0.00000000000000  B
   0.80000000000000   0.60000000000000   0.00000000000000  B
   0.40000000000000   0.80000000000000   0.00000000000000  B
```

Using the default tolerance for all routines, Minkowski transformation and reducing the lattice modifies the basis: (1) $a' \rightarrow a - 2c + b$; (2) $b' \rightarrow b - c$

```
WARNING: using default tolerance 0.001
Using tolerance: 0.001
MINKOWSKI-REDUCED PRIMITIVE CELL
AB/285 - (285) - ???? [AB] (285)  (htqc library)
-123432.463200
   0.00000000000000   0.00000000000000  -3.56400000000000
  -2.86000000000000   2.86000000000000  -1.78200000000000
   2.86000000000000   2.86000000000000   1.78200000000000
1 4
Direct(5) [A1B4]
   0.00000000000000   0.00000000000000   0.00000000000000  A
   0.20000000000000   0.20000000000000   0.60000000000000  B
   0.60000000000000   0.60000000000000   0.80000000000000  B
   0.80000000000000   0.80000000000000   0.40000000000000  B
```

```
0.40000000000000   0.40000000000000   0.20000000000000  B
```

The orthogonality defects can be compared: The transformed representation has $\delta = 1.19$ and the original representation has $\delta = 2.48$. The volume remains the same because the original representation is primitive.

Analysis of the lattice reveals five mirror operations; thus, the crystal is conformed to tetragonal symmetry. A complete search for symmetry elements using the symmetry library gives the following:

```
latticetypechar: t
symlist: "1 1 1 0 2 0 0 0 2 0 "
Original 'crystal system' / lattice system: TETRAGONAL
CRYSTAL SYSTEM: TETRAGONAL
POINT GROUP: 4/m
```

Using the known symmetry properties, the standard conventional cell is constructed:

```
CONVENTIONAL CELL CHOICE:
AB/285 - (285) - ???? [AB] (285)
-246865
5.7200000000000 0.0000000000000 0.0000000000000
0.0000000000000 5.7200000000000 0.0000000000000
0.0000000000000 0.0000000000000 3.5640000000000
8 2
Direct(10)
0.10000000000000 0.70000000000000 0.50000000000000 1
0.70000000000000 0.90000000000000 0.50000000000000 1
0.20000000000000 0.40000000000000 0.0000000000000 1
0.60000000000000 0.20000000000000 0.0000000000000 1
```

```
0.30000000000000 0.10000000000000 0.50000000000000 1
0.40000000000000 0.80000000000000 0.0000000000000 1
0.90000000000000 0.30000000000000 0.50000000000000 1
0.80000000000000 0.60000000000000 0.0000000000000 1
0.50000000000000 0.50000000000000 0.50000000000000 0
0.0000000000000 0.0000000000000 0.0000000000000 0
```

The full symmetry set is known:

```
1          0.0000e+00  0.0000e+00  0.0000e+00
1          5.0000e-01  5.0000e-01  5.0000e-01
-1 0,0,0    5.0000e-01  5.0000e-01  5.0000e-01
-1 0,0,0    0.0000e+00  0.0000e+00  0.0000e+00
2 0,0,z    5.0000e-01  5.0000e-01  5.0000e-01
2 0,0,z    0.0000e+00  0.0000e+00  0.0000e+00
4+ 0,0,z    0.0000e+00  0.0000e+00  0.0000e+00
4+ 0,0,z    5.0000e-01  5.0000e-01  5.0000e-01
4- 0,0,z    0.0000e+00  0.0000e+00  0.0000e+00
4- 0,0,z    5.0000e-01  5.0000e-01  5.0000e-01
m x,y,0    0.0000e+00  0.0000e+00  0.0000e+00
m x,y,0    5.0000e-01  5.0000e-01  5.0000e-01
-4+ 0,0,z    0.0000e+00  0.0000e+00  0.0000e+00
-4+ 0,0,z    5.0000e-01  5.0000e-01  5.0000e-01
-4- 0,0,z    0.0000e+00  0.0000e+00  0.0000e+00
-4- 0,0,z    5.0000e-01  5.0000e-01  5.0000e-01
```

Given the Bravais type of body-centered (I) and the point group $(4/m)$ the candidate space group set is reduced to one of two: #87-#88. Searching the generator sets for a match leaves only space group #87.

The set of standard Wyckoff positions (from the International Tables) is loaded for space group #87:

```
Space Group 87
 (0,0,0) (.5,.5,.5)
 16 i 1 (x,y,z) (-x,-y,z) (-y,x,z) (y,-x,z)
 (-x,-y,-z) (x,y,-z) (y,-x,-z) (-y,x,-z)
 8 h m.. (x,y,0) (-x,-y,0) (-y,x,0) (y,-x,0)
 8 g 2.. (0,.5,z) (.5,0,z) (0,.5,-z) (.5,0,-z)
 8 f -1 (.25,.25,.25) (.75,.75,.25) (.75,.25,.25) (.25,.75,.25)
 4 e 4.. (0,0,z) (0,0,-z)
 4 d -4.. (0,.5,.25) (.5,0,.25)
 4 c 2/m.. (0,.5,0) (.5,0,0)
 2 b 4/m.. (0,0,.5)
 2 a 4/m.. (0,0,0)
```

Separating the atomic basis into equivalent atom groupings and then solving for the Wyckoff variables gives two positions: 'a' and 'h'. The Wyckoff variables as a triple (x,y,z), labels, and site symmetry are then given.

```
6.0000e-01  2.0000e-01  0.0000e+00  B  8 h m..
0.0000e+00  0.0000e+00  0.0000e+00  A  2 a 4/m..
```

Finally, a structure representation using the the conventional cell and the Wyckoff positions (the "WYCCAR") can be displayed (space group number, point group , and Bravais label are included in the title):

```
[TITLE] | SG #87 PG: 4/m BL: I
-246865
 5.7200e+00  3.5025e-16 -0.0000e+00
```

```
-3.5025e-16  5.7200e+00  0.0000e+00
 0.0000e+00  0.0000e+00  3.5640e+00
8 2
DIRECT(WYCCAR)
 6.0000e-01  2.0000e-01  0.0000e+00 B 8 h m..
 0.0000e+00  0.0000e+00  0.0000e+00 A 2 a 4/m..
```

4.7 Performance of the space group routine

The performance of the space group algorithm is determined primarily by the full symmetry element routine and the origin shift/generator match routine. The scaling of the topological symmetry element search in the conventional cell routine depends only on the lattice and so is fixed with respect to the number of atoms in the cell. The lattice reduction and primitivization routines are well-studied (Buchmann (2004); Santoro and Mighell (1970); Wübben et al. (2011); Grosse-Kunstleve et al. (2004)) and are minor contributors to the overall routine's scaling with atoms in the basis (Fig. 4.2).

The full-symmetry routine loads the point operator matrices $\{R\}$ that are relevant to the known crystal system. For cubic crystals there are forty-eight matrices; for hexagonal crystals there are twenty-four; for tetragonal crystals there are sixteen; for orthorhombic crystals there are eight; for monoclinic crystals there are four; and for triclinic crystals there are two. For each of the relevant point operators, the smallest set of chemically equivalent atoms is subtracted from a fixed point in the same set to define a candidate internal translation. So, for a primitive cell with N atoms there are at most $a \leqslant N$ candidate internal translations that must be applied with each candidate point operator to the atomic basis. This requires at most an additional N operations to confirm or reject that the $\{R, t\}$ pair is an isometry of the crystal. If we denote the number of operators in $\{R\}$ as m, at worst this routine scales like

$\mathcal{O}(mN^2)$. In practice, the routine tends to do much better, as an operator $\{R, t\}$ will often be rejected by operation on an atom before reaching the end of the basis.

The search for the generator set and origin shift does not depend on the number of atoms in the unit cell but depends on the number of generators for the space group and the centering operations of the lattice. For a space group with G generators and C centered sets, there are C^G groupings of symmetry elements that must be tested for an origin shift. In the worst case of FCC cubic, $G = 5$ and $C = 4$; thus, there are $4^5 = 1024$ linear problems that must be checked for solution. The linear problem is solution of an over-determined system which scales like $\mathcal{O}(G^2)$.

The Wyckoff position routine scales with the number of atoms and the size of the symmetry set at worst $\mathcal{O}(mN)$. This arises from first grouping the atoms by symmetrically equivalent types. For an atomic basis $a_1, a_2, ..., a_N$ and symmetry element set $R_1, R_2, ..., R_m$, the routine begins by placing a_1 and all sites derived by application of the symmetry set into the first bin. The same is done for a_2 and all its derivatives, for which a new bin will be created if it is not contained within the first bin. This is carried out for all atoms. One could improve this routine by dynamically updating the set of atoms that are checked to exclude those that are derived, but the gain in efficiency is small relative to the dominating scaling of lattice reduction and the full-symmetry routine.

4.8 Reverse Space Group Function

In order to test and debug the symmetry code, a "reverse" space group function was written. This utilizes the libraries containing the lattice and atomic basis data in the International Tables. When the space group number, Wyckoff multiplicity group, and index for the given multiplicity group are provided, a crystal structure in the POSCAR format is constructed with the specified space group and Wyckoff site. The parameters for Wyckoff positions with at least one parametric degree of freedom

are selected automatically and uniquely. The routine can also be simply modified to select pseudo-randomly generated parameters or from user-defined inputs. The reverse space group function was used to test the scaling of the code with increasing atomic basis size *within* the same space group and to debug the software.

4.9 Tolerance Control

Finally, the role of tolerance and accuracy in symmetry computations is significant. It should be emphasized that there is no "ideal" tolerance setting (ie, the ideal tolerance is problem and structure dependent). As the tolerance factors used within the many functions that comprise the space group routine are varied, symmetry elements may appear or be extinguished, thus altering the computed space group number and related symmetry properties. Thus, it is the obligation of the user to think carefully about the structural information that is provided to the code as well as the relevance of the computed output.

With the problem-specific nature of tolerance in the context of computational symmetry routines, the code has been written to easily allow for modifications of the tolerance factors within the various functions. The globally-defined tolerance factors can be controlled individually within the `void SetTolerance()` function for fine-tuning, or to define the value of all simultaneously, the user can provide a tolerance value on the command line execution of the code.

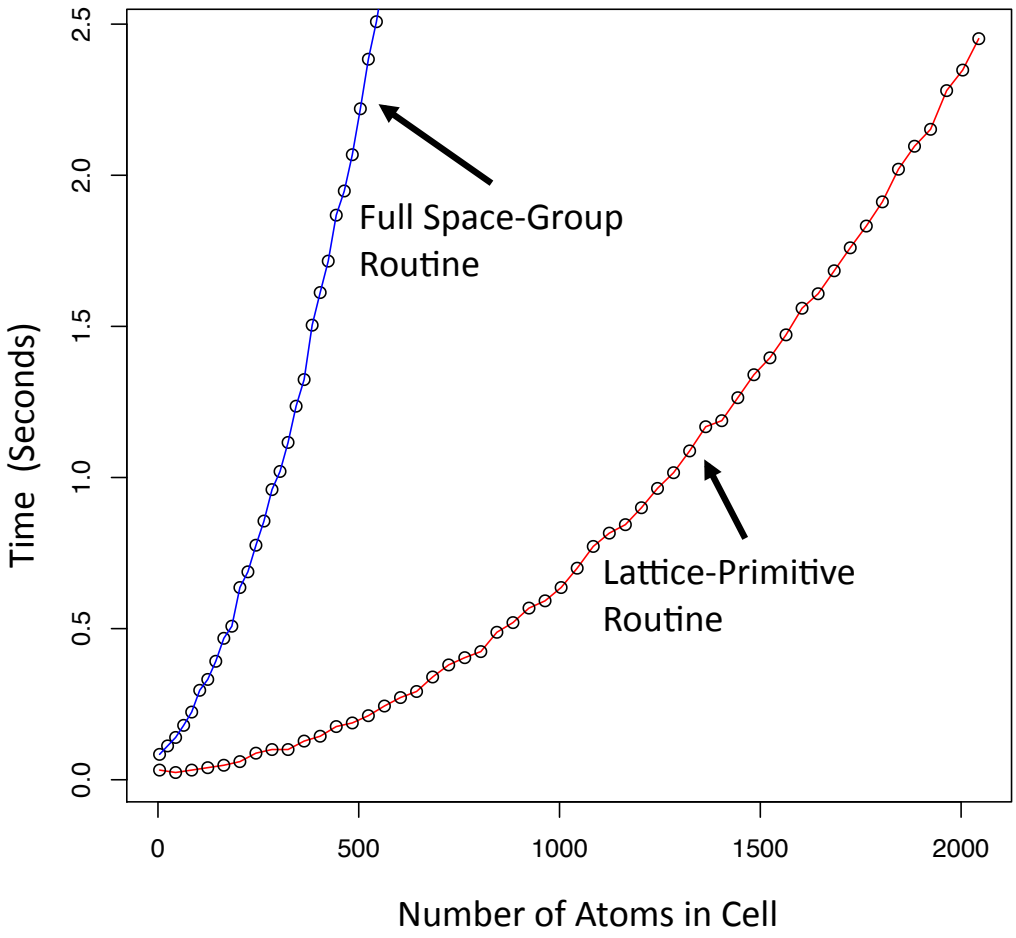


FIGURE 4.2: Representative scaling of the lattice primitive routine and the full space group routine (including the primitive cell routine) with number of atom in cell. The cells analyzed were "reverse" constructed crystals with space group number 11 and atoms in the general Wyckoff position. The qualitative behavior of the scaling applies to all space group types.

5

Summary and Conclusions

In conclusion, software was written in C++ for the full symmetry characterization of a general crystal structure. The space group, point group, Bravais lattice, crystal system, and lattice system are determined and the structure is represented in a conventional form with a standard origin according to the International Tables of Crystallography (ITC) (Hahn (2002)). Furthermore, the Wyckoff positions are determined in the standard form and a reduced representation in terms of the representative symmetry positions is produced (the Wyccar).

The software first performs lattice reduction and primitivization to account for skew cells and possible non-primitive and supercell representations. The primitive, Minkowski-reduced cell is then subject to a topologically-based lattice analysis which is used to construct the first iteration of the standard conventional cell. The standard conventional cell is one which makes use of the symmetry elements to select and orient the lattice vectors. The group-subgroup relations of the thirty-two crystal classes (crystallographic point groups) allow for a systematic progression toward the correct conventional cell by utilizing a library of crystallographically permissible point operations. This occurs after at most one iteration. Additional iterations of the

ordering of the lattice vectors may be necessary when equivalent symmetry directions are used to define the lattice vectors (eg, in orthorhombic cells).

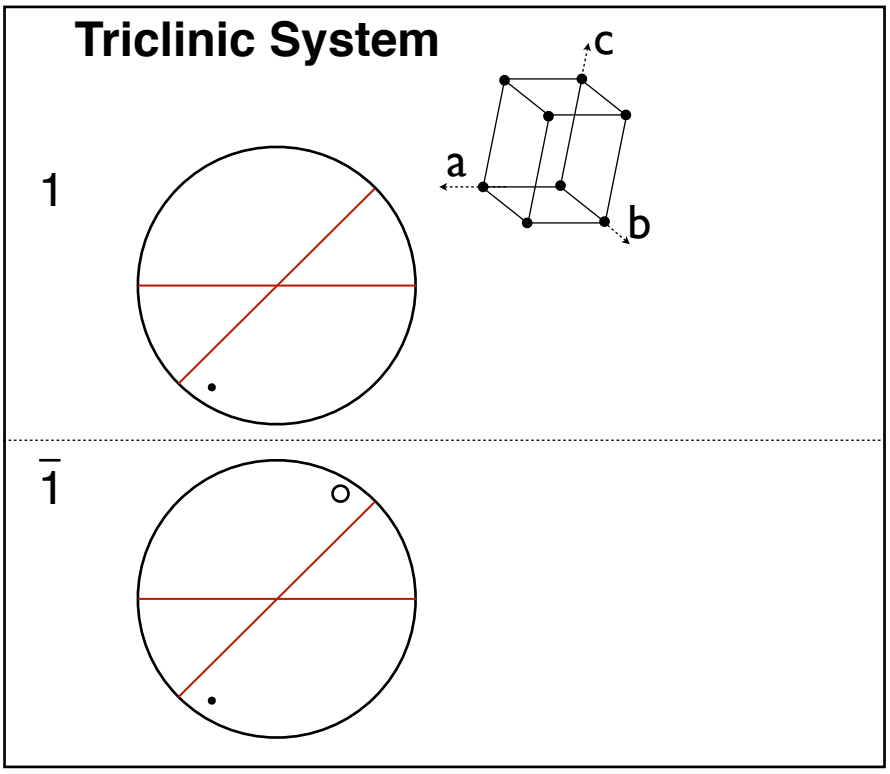
Once the conventional cell has been determined, the point group and Bravias lattice are simultaneously known, and the candidate space group set is dramatically reduced from the existing 230 possibilities. The sets of space group generators within the reduced set of candidate space groups are searched for a match with the input structure. This provides not only the space group but also the origin shift to the standard (ITC) representation.

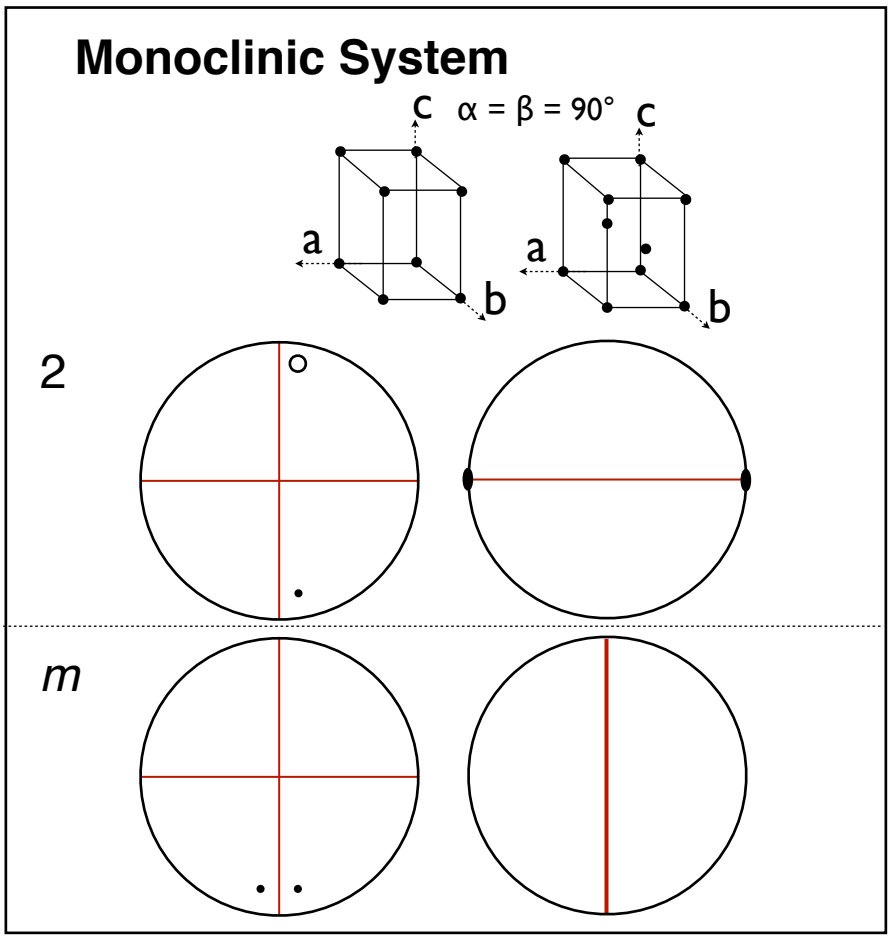
The origin shift is then applied to the structure in question, and the atomic basis is partitioned using the full symmetry set into symmetrically equivalent atoms. The partitioned set of atoms limits the possible Wyckoff sites but does not uniquely define the position or provide the value of the site parameters. This problem is solved by using a library of Wyckoff positions for the 230 space groups, taken from the ITC. The Wyckoff sites with degeneracies matching the degeneracies of the sets of equivalent atoms are equated with the equivalent atoms and a consistent set of parameters is sought. Those with consistency represent Wyckoff positions of the original structure. The Wyckoff sites symmetries, labels, and parameter values are compiled into the Wyccar structure representation.

The software has been fully implemented within AFLOW – a high-throughput computational framework for materials discovery and design. The underlying structure class, `xstructure`, has been used to optimize future use and augmentation. Also, because the software has been written in object-oriented C++ and makes use of the underlying crystal structure classes within AFLOW future use, extensions, and modifications, will be greatly facilitated.

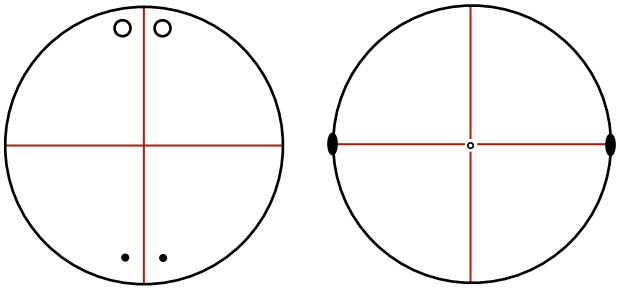
Appendix A

Stereographic projections of the 32 point groups including crystal system and the Bravais lattices



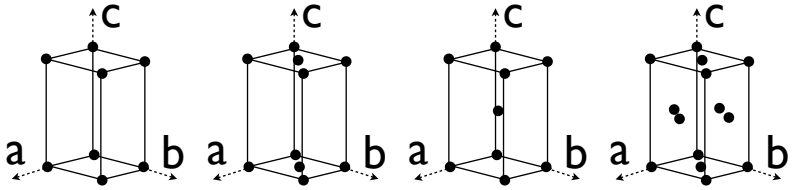

Monoclinic System
c
α = β = 90°
c
a
b
a
b
2
m

2/*m*

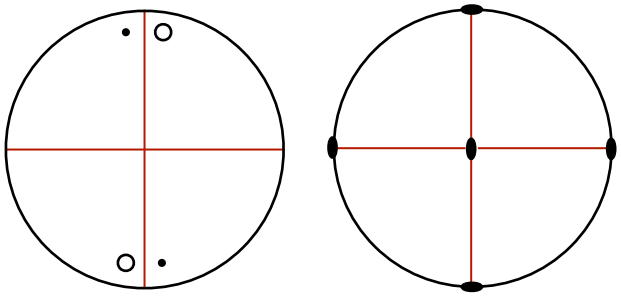


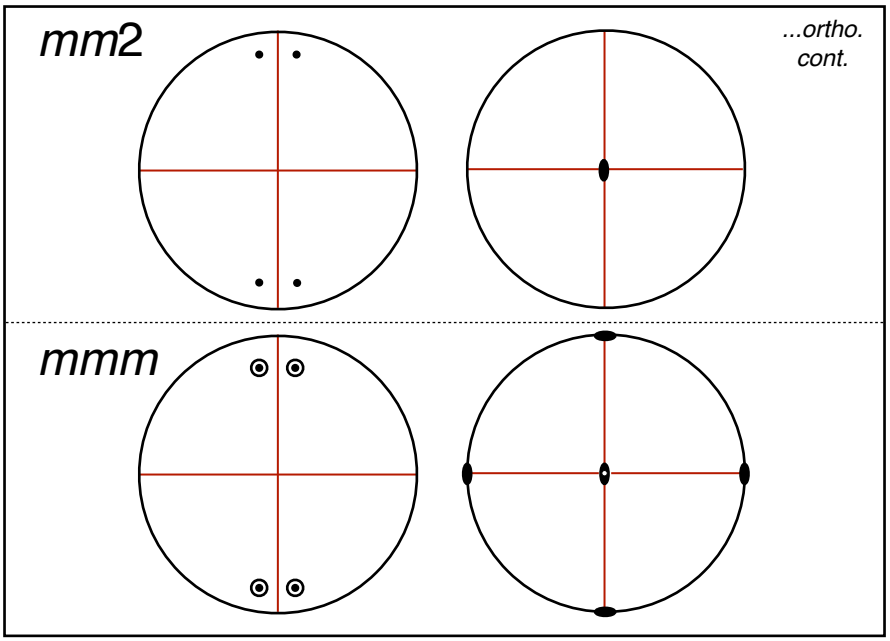
... mono. cont.

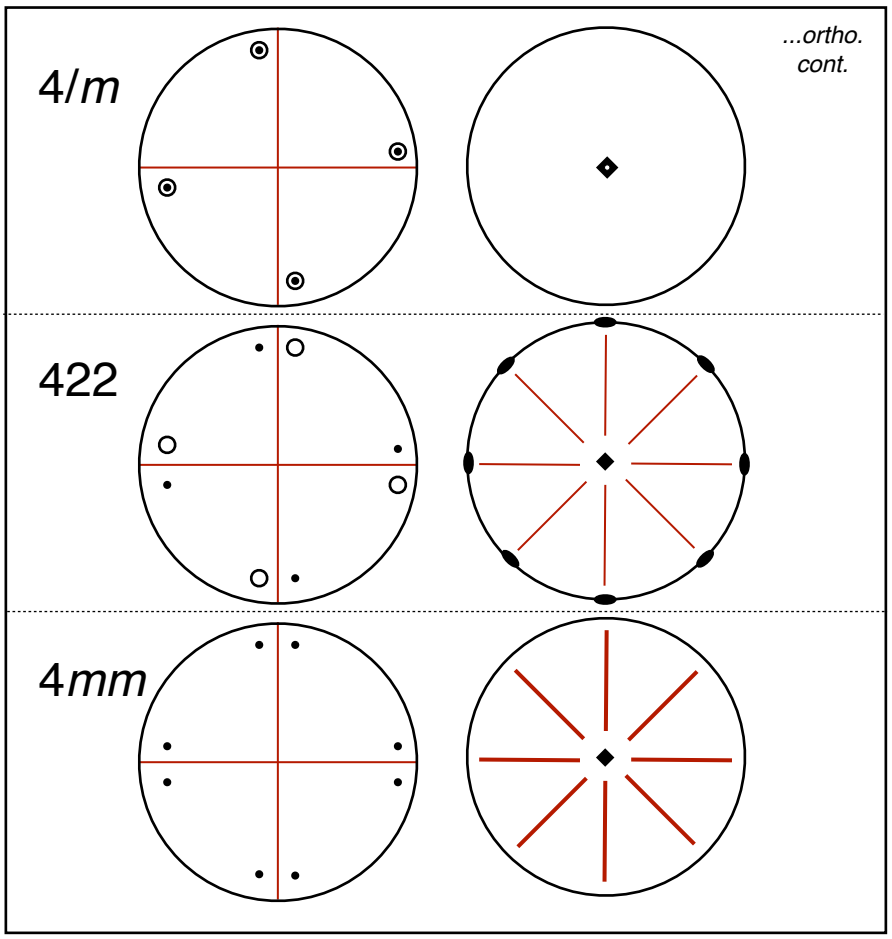
Orthorhombic System $\alpha = \beta = \gamma = 90°$

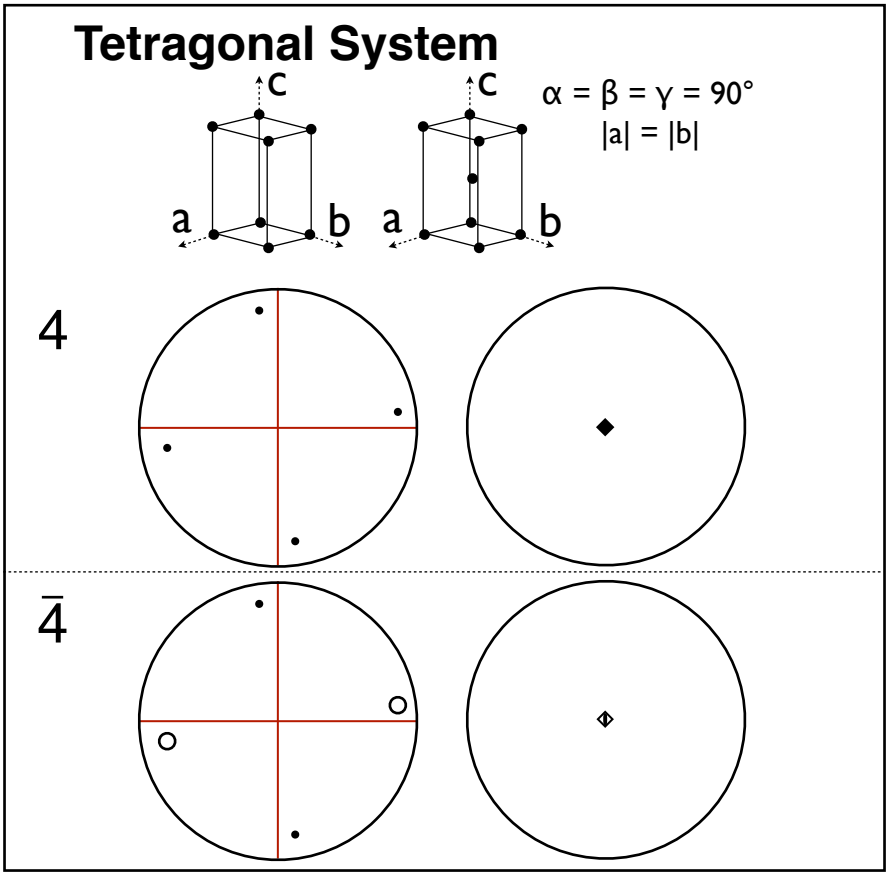


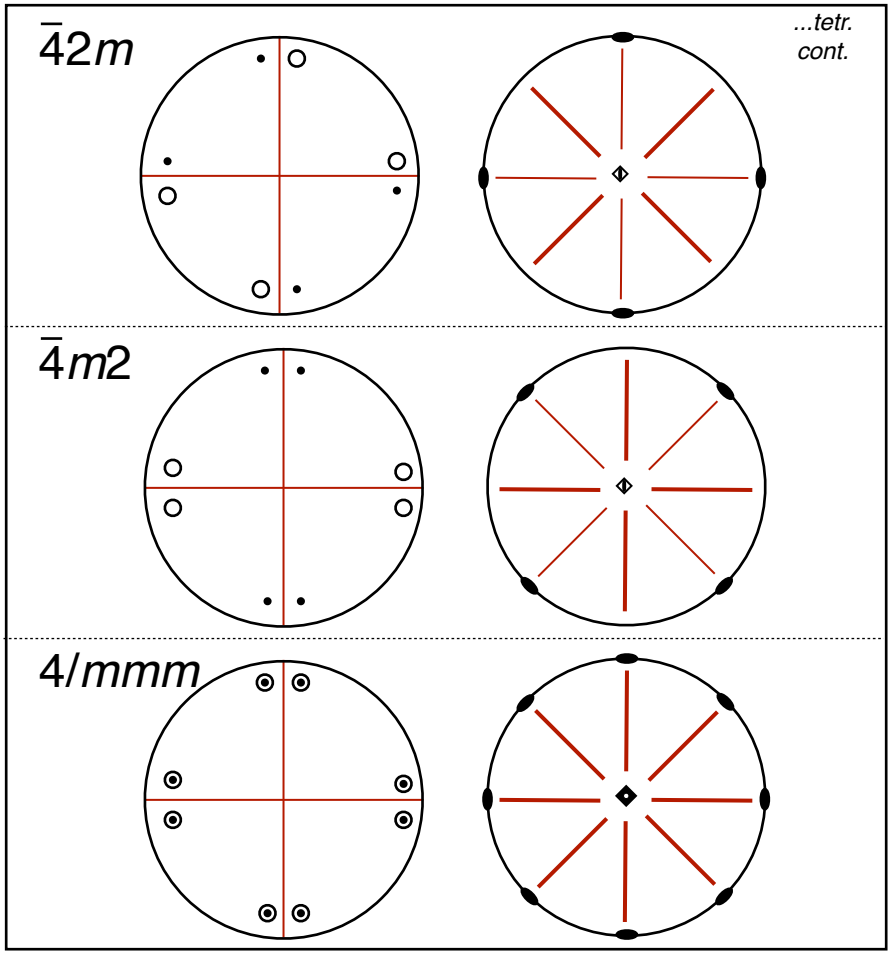
222

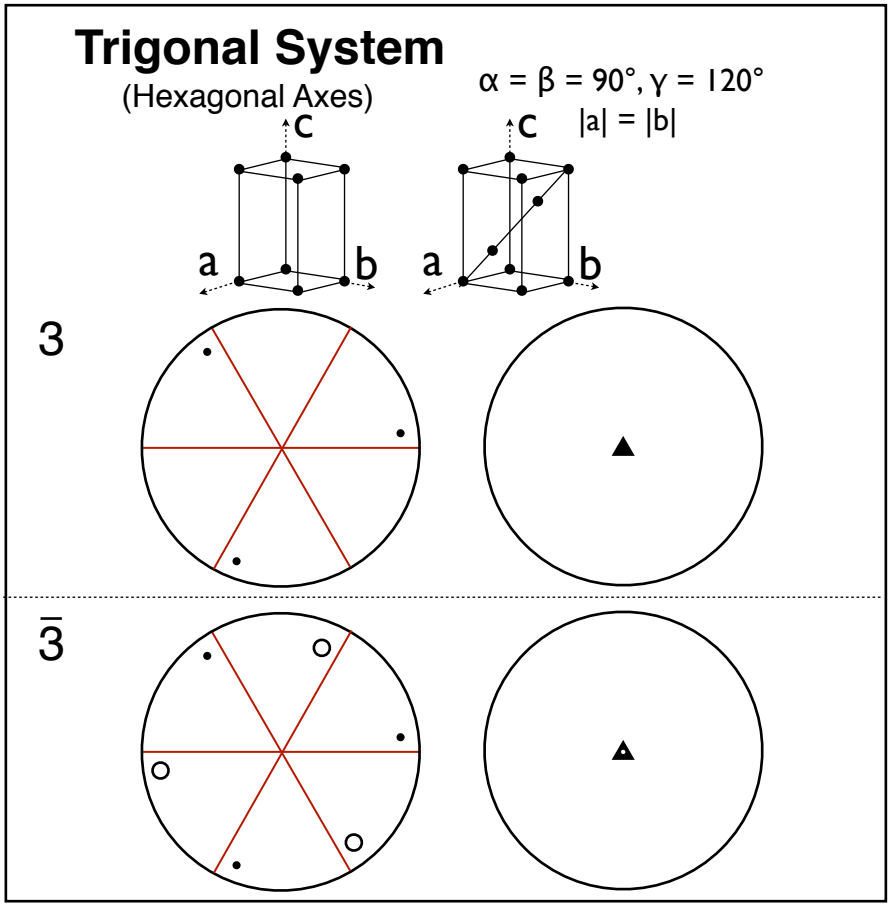


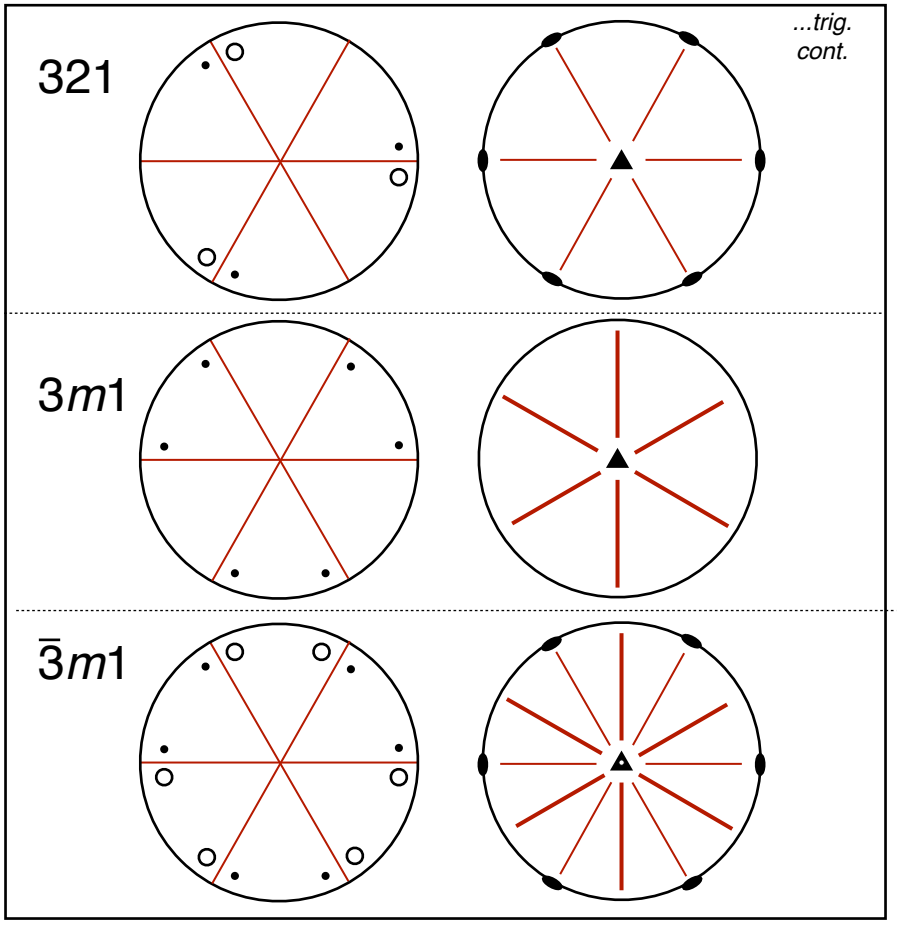

mm2
...ortho.
cont.
mmm


...ortho.
cont.
4/m
422
4mm

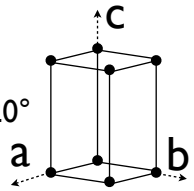

Tetragonal System
c
a
b
c
a
b
α = β = γ = 90°
|a| = |b|
4
4̄


...tetr.
cont.
$\bar{4}2m$
$\bar{4}m2$
4/mmm


Trigonal System
(Hexagonal Axes)
α = β = 90°, γ = 120°
|a| = |b|
c
a
b
c
a
b
3
3̄


...trig.
cont.
321
3m1
3̄m1

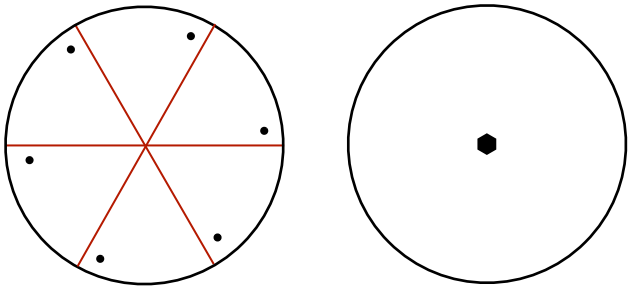
Hexagonal System



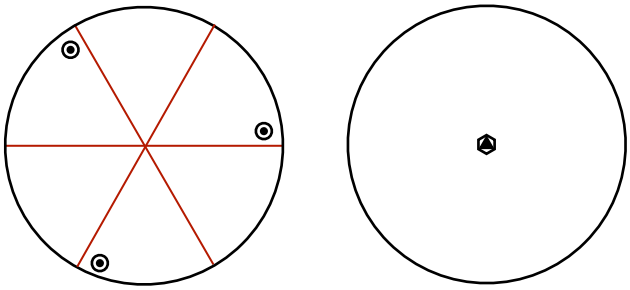
$\alpha = \beta = 90°, \gamma = 120°$

$|a| = |b|$

6

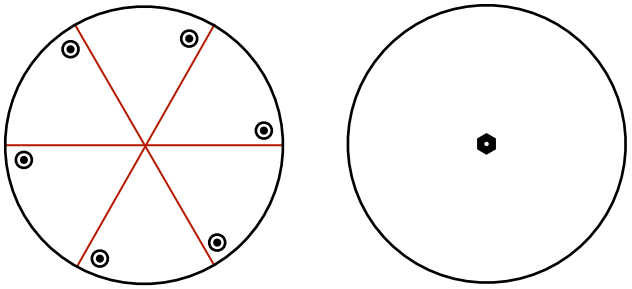


$\bar{6}$

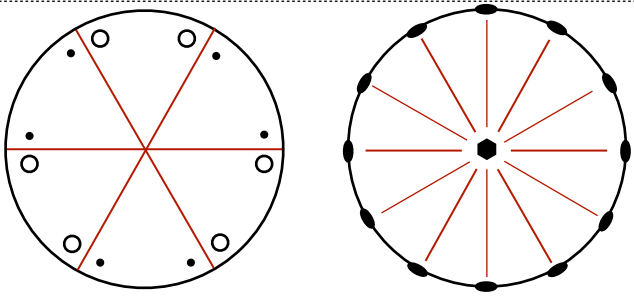


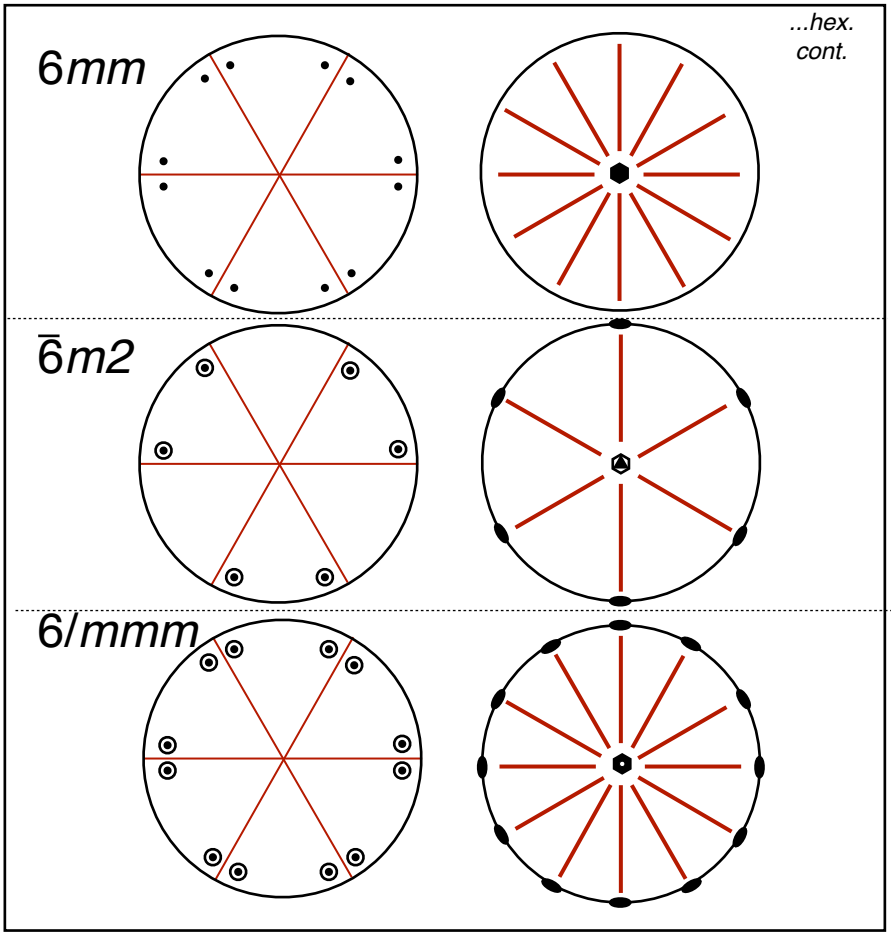
...hex. cont.

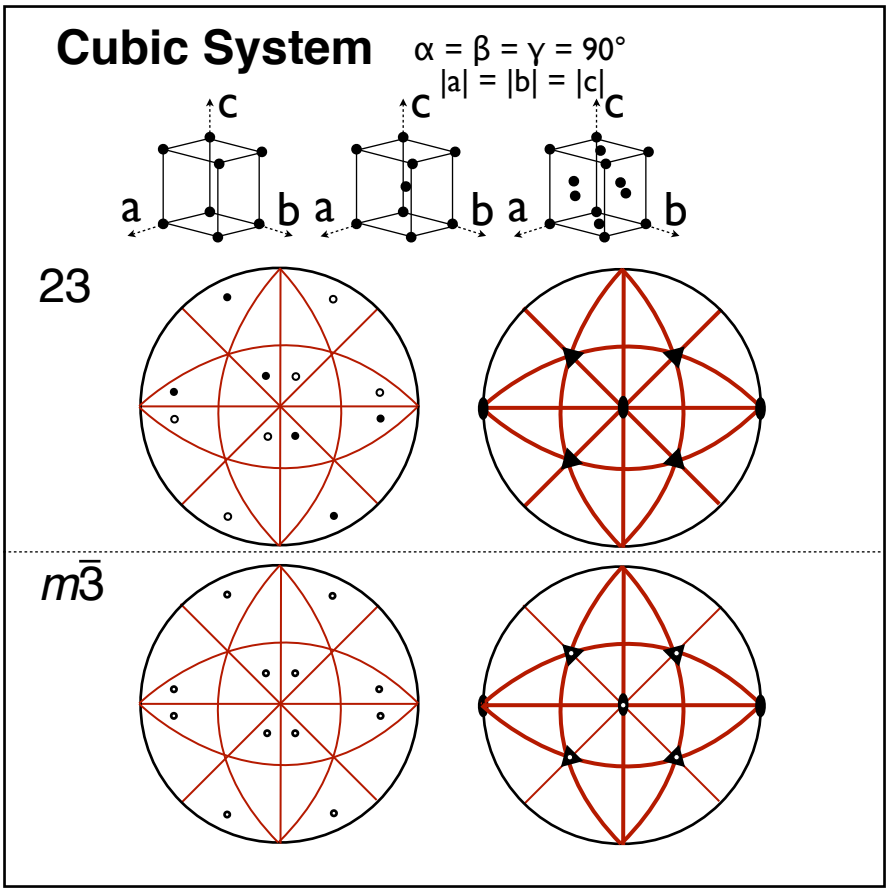
6/*m*



622




...hex.
cont.
6mm
6̄m2
6/mmm


Cubic System
α = β = γ = 90°
|a| = |b| = |c|
c
a
b
23
m3̄

Bibliography

Buchmann, A. (ed.) (2004), *Lecture Notes in Computer Science*, Springer.

Curtarolo, S., Setyawan, W., Hart, G. L. W., Jahnatek, M., Chepulskii, R. V., Taylor, R. H., Wang, S., Xue, J., Yang, K., Levy, O., Mehl, M., Stokes, H. T., Demchenko, D. O., and Morgan, D. (2012a), "AFLOW: an automatic framework for high-throughput materials discovery," *AFLOW: an automatic framework for high-throughput materials discovery, submitted to Comp. Mat. Sci.*

Curtarolo, S., Setyawan, W., Wang, S., Xue, J., Yang, K., Taylor, R. H., Nelson, L. J., Hart, G. L. W., Sanvito, S., Buongiorno Nard elli, M., Mingo, N., and Levy, O. (2012b), "AFLOWLIB.ORG: A distributed materials properties repository from high-throughput *ab initio* calculations," *Comp. Mat. Sci.*, 58, 227–235.

Curtarolo, S., Hart, G. L. W., Nardelli, M. B., Mingo, N., Sanvito, S., and Levy, O. (2013), "The high-throughput highway to computational materials design," *Nature Materials*, 12, 191–201.

Donnay, J. D. H. (1943), "Rules for the conventional orientation of crystals," *American Mineralogist*, 28, 313–328.

Eisenstein, G. (1851)*J. Math. (Crelle)*, 41, 141.

Grosse-Kunstleve, R. W., Sauter, N. K., and Adams, P. D. (2004), "Numerically stable algorithms for the computation of reduced unit cells," *Acta Crystallographica Section A*, 60, 1–6.

Hahn, T. (ed.) (2002), *International Tables of Crystallography. Volume A: Space-group symmetry*, Kluwer Academic Publishers, International Union of Crystallography, Chester, England.

Kresse, G. and Furthmuller, J. (1996), "Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set," *Comp. Mat. Sci.*, 6, 15.

Kresse, G. and Hafner, J. (1993), "Ab initio molecular dynamics for liquid metals," *Phys. Rev. B*, 47, 558–561.

Niggli, P. (1928), *Handbuch der Experimentalphysik*, vol. 7, Akademische Verlagsgesellschaft.

S. Curtarolo, W. Setyawan, G. L. W. Hart, M. Jahnatek, R. V. Chepulskii, R. H. Taylor, S. Wang, J. Xue, K. Yang, O. Levy, M. Mehl, H. T. Stokes, D. O. Demchenko, and D. Morgan (2012), "AFLOW: an automatic framework for high-throughput materials discovery," *AFLOW: an automatic framework for high-throughput materials discovery, submitted to Comp. Mat. Sci.*

Santoro, A. and Mighell, A. D. (1970), "Determination of reduced cells," *Acta Crystallographica Section A*, 26, 124–127.

Stokes, H. T., Campbell, B. J., and Hatch, H. T. (2013), "FINDSYM," .

Wübben, D., Seethaler, D., Jalendn, J., and Matz, G. (2011), "Lattice Reduction," *Signal Processing Magazine, IEEE*, 28, 70 –91.

Biography

- Richard Hansen Taylor II
- November 1, 1985, La Jolla San Diego, California
- M.S., Mechanical Engineering, Duke University; M.S. Physics, Brigham Young University, B.S. Physics, Brigham Young University

Publications

- S. Curtarolo, W. Setyawan, S. Wang, J. Xue, K. Yang, **R. H. Taylor**, G. L. W. Hart, S. Sanvito, M. Buongiorno-Nardelli, N. Mingo, and O. Levy, AFLOWLIB.ORG: a distributed materials properties repository from high-throughput ab initio calculations, in press, Comp. Mat. Sci. (2012).
- S. Curtarolo, W. Setyawan, G. L. W. Hart, M. Jahnatek, R. V. Chepulskii, **R. H. Taylor**, S. Wang, J. Xue, K. Yang, O. Levy, M. Mehl, H. T. Stokes, D. O. Demchenko, and D. Morgan, AFLOW: an automatic framework for high-throughput materials discovery, in press, Comp. Mat. Sci. (2012).
- **R. H. Taylor**, S. Curtarolo, and G. L. W. Hart, Guiding the experimental discovery of magnesium alloys, Phys. Rev. B 84, 084101 (2011).
- **R. H. Taylor**, S. Curtarolo, and G. L. W. Hart, Predictions of the Pt8Ti phase in unexpected systems, J. Am. Chem. Soc. 132, 6851-6854 (2010).
- **R. H. Taylor**, S. Curtarolo, and G. L. W. Hart, Ordered Magnesium-Lithium Alloys: First-Principles Predictions, Phys. Rev. B 81, 024112 (2010).

Awards/Fellowships

Pratt-Gardner Fellow (2011) NIST Guest Researcher (2012-2013)